

TEHNIČKI GLASNIK - TECHNICAL JOURNAL

Scientific-professional journal of University North

Volume 13

Varaždin, December 2019

Number 4

Pages 261–362

Editorial Office:

Sveučilište Sjever – Tehnički glasnik
 Sveučilišni centar Varaždin
 104. brigade 3, 42000 Varaždin, Hrvatska
 Tel. ++385 42 493 328, Fax. ++385 42 493 333
 E-mail: tehnickiglasnik@unin.hr
<https://tehnickiglasnik.unin.hr>
<https://www.unin.hr/djelatnost/izdavastvo/tehnicki-glasnik/>
<https://hrcak.srce.hr/tehnickiglasnik>

Founder and Publisher:

Sveučilište Sjever / University North

Council of Journal:

Marin MILKOVIĆ, Chairman; Anica HUNJET, Member; Goran KOZINA, Member; Mario TOMIŠA, Member;
 Vlado TROPŠA, Member; Damir VUSIĆ, Member; Milan KLJAJIN, Member; Anatolii KOVROV, Member

Editorial Board:

Chairman Damir VUSIĆ (1), Milan KLJAJIN (2)/(1), Marin MILKOVIĆ (1), Krešimir BUNTAK (1), Anica HUNJET (1), Živko KONDIĆ (1), Goran KOZINA (1), Ljudevit KRPAN (1), Krunoslav HAJDEK (1), Marko STOJIC (1), Božo SOLDIĆ (1), Mario TOMIŠA (1), Vlado TROPŠA (1), Vinko VIŠNJIĆ (1), Duško PAVLETIĆ (5), Branimir PAVKOVIĆ (5), Mile MATIJEVIĆ (3), Damir MODRIĆ (3), Nikola MRVAC (3), Klaudio PAP (3), Ivana ŽILJAK STANIMIROVIĆ (3), Krešimir GRILEC (6), Biserka RUNJE (6), Predrag ČOSIĆ (6), Sara HAVRLIŠAN (2), Dražan KOZAK (2), Roberto LUJIĆ (2), Leon MAGLIĆ (2), Ivan SAMARDŽIĆ (2), Antun STOIC (2), Katica ŠIMUNOVIĆ (2), Goran ŠIMUNOVIĆ (2), Ladislav LAZIĆ (7), Ante ČIKIĆ (1)/(2), Darko DUKIĆ (9), Gordana DUKIĆ (10), Srdan MEDIC (11), Sanja KALAMBURA (12), Marko DUNĐER (13), Zlata DOLACEK-ALDUK (4), Dina STOBBER (4)

International Editorial Council:

Boris TOVORNIK (14), Milan KUHTA (15), Nenad INJAC (16), Džafer KUDUMOVIC (17), Marin PETROVIĆ (18), Salim IBRAHIMEFENDIĆ (19), Zoran LOVREKOVIĆ (20), Igor BUDAK (21), Darko BAJIĆ (22), Tomáš HANÁK (23), Evgenij KLIMENKO (24), Oleg POPOV (24), Ivo ČOLAK (25), Katarina MONKOVA (26), Berenika HAUSNEROVÁ (8)

Editor-in-Chief:

Milan KLJAJIN

Technical Editor:

Goran KOZINA

Graphics Editor:

Snježana IVANČIĆ VALENKO

Linguistic Advisers for English language:

Ivana GRABAR, Iva GRUBJEŠIĆ

IT support:

Tomislav HORVAT

Print:

Centar za digitalno nakladništvo, Sveučilište Sjever

All manuscripts published in journal have been reviewed.**Manuscripts are not returned.****The journal is free of charge and four issues per year are published.****Circulation:** 100 copies**Journal is indexed and abstracted in:**

Web of Science Core Collection (Emerging Sources Citation Index - ESCI), EBSCOhost Academic Search Complete, EBSCOhost – One Belt, One Road Reference Source Product, ERIH PLUS, CITEFACTOR – Academic Scientific Journals, Hrcak - Portal znanstvenih časopisa RH

Registration of journal:

The journal "Tehnički glasnik" is listed in the HGK Register on the issuance and distribution of printed editions on the 18th October 2007 under number 825.

Preparation ended:

December 2019

Legend:

(1) University North, (2) Mechanical Engineering Faculty in Slavonski Brod, (3) Faculty of Graphic Arts Zagreb, (4) Faculty of Civil Engineering Osijek, (5) Faculty of Engineering Rijeka, (6) Faculty of Mechanical Engineering and Naval Architecture Zagreb, (7) Faculty of Metallurgy Sisak, (8) Tomas Bata University in Zlin, (9) Department of Physics of the University of Josip Juraj Strossmayer in Osijek, (10) Faculty of Humanities and Social Sciences Osijek, (11) Karlovac University of Applied Sciences, (12) University of Applied Sciences Velika Gorica, (13) Department of Polytechnics - Faculty of Humanities and Social Sciences Rijeka, (14) Faculty of Electrical Engineering and Computer Science - University of Maribor, (15) Faculty of Civil Engineering - University of Maribor, (16) University College of Teacher Education of Christian Churches Vienna/Krems, (17) Mechanical Engineering Faculty Tuzla, (18) Mechanical Engineering Faculty Sarajevo, (19) University of Travnik - Faculty of Technical Studies, (20) Higher Education Technical School of Professional Studies in Novi Sad, (21) University of Novi Sad - Faculty of Technical Sciences, (22) Faculty of Mechanical Engineering - University of Montenegro, (23) Brno University of Technology, (24) Odessa State Academy of Civil Engineering and Architecture, (25) Faculty of Civil Engineering - University of Mostar, (26) Faculty of Manufacturing Technologies with the seat in Prešov - Technical University in Košice

CONTENT	I
NOTE FROM THE EDITOR-IN-CHIEF	II
Hnin Thiri CHAW, Sinchai KAMOLPHIWONG, Krongthong WONGSRITRANG SLEEP APNEA DETECTION USING DEEP LEARNING	261
Aziz NANTHAAMORNPHONG, Stephane BRESSAN THE EMPIRICAL STUDY: ENCOURAGING STUDENTS' INTEREST IN SOFTWARE DEVELOPMENT USING TEST-DRIVEN DEVELOPMENT	267
Sumalee SANGAMUANG A GRAPH-BASED ALGORITHM FOR INTERPERSONAL TIES CLUSTERING IN SIGNED NETWORKS	275
Ei Ei MON, Sangsuree VASUPONGAYYA, Montri KARNJANADECHA, Touchai ANGCHUAN EVALUATING BIOMETRICS FINGERPRINT TEMPLATE PROTECTION FOR AN EMERGENCY SITUATION	280
Siraphob THEERACHEEP, Jaruloj CHONGSTITVATANA MULTIPLICATION OF MEDIUM-DENSITY MATRICES USING TENSORFLOW ON MULTICORE CPUs	286
Preecha VONGHIRANDECHA, Montri KARNJANADECHA, Sathit INTAJAG CONTRAST AND COLOR BALANCE ENHANCEMENT FOR NON-UNIFORM ILLUMINATION RETINAL IMAGES	291
Memduha ERGÜT, Ayla ÖZER HETEROGENEOUS FENTON-LIKE DECOLORIZATION OF PROCION RED MX-5B WITH IRON-ALGINATE GEL BEADS AS AN EFFECTIVE CATALYST	297
Marko MORIĆ, Igor MAJNARIĆ, Klaudio PAP, Slaven MILOŠ THE INFLUENCE OF PRE-TREATMENT BY PRIMING ON THE CMY REPRODUCTION QUALITY PRINTED WITH ELECTROINK	305
Vilko ŽILJAK, Jana ŽILJAK GRŠIĆ, Denis JUREČIĆ, Tonči JELIČIĆ NEAR-INFRARED SPECTROSCOPY AND HIDDEN GRAPHICS APPLIED IN PRINTING SECURITY DOCUMENTS IN THE OFFSET TECHNIQUE	311
Željko KNOK, Klaudio PAP, Marko HRNČIĆ IMPLEMENTATION OF INTELLIGENT MODEL FOR PNEUMONIA DETECTION	315
Alireza SINA, Damanjeet KAUR LOAD FREQUENCY CONTROL OF MULTI AREA INTERCONNECTED POWER SYSTEM USING DIFFERENTIAL EVOLUTION ALGORITHM	323
Zeljko KOS, Valentyn PARUTA, Olga GNYP, Leonid LAVRENYUK, Iryna GRYNYOVA DESIGNING PLASTER MORTAR FOR A HIGH-RISE BUILDING	331
Ercan AVŞAR, Kerem SALÇIN DETECTION AND CLASSIFICATION OF BRAIN TUMOURS FROM MRI IMAGES USING FASTER R-CNN	337
Adam GNATOWSKI, Rafał GOŁĘBSKI, Piotr SIKORA ANALYSIS OF THERMOMECHANICAL PROPERTIES OF POLYMERIC MATERIALS PRODUCED BY A 3D PRINTING METHOD	343
Jiri TUPA, Frantisek STEINER INDUSTRY 4.0 AND BUSINESS PROCESS MANAGEMENT	349
Jerko GLAVAŠ, Veljko KONDIĆ, Iva BALIĆ LEADERSHIP AND CONFLICT MANAGEMENT IN PRODUCTION ORGANISATIONS	356
INSTRUCTIONS FOR AUTHORS	III



Note from the Editor-in-Chief

Dear readers,

The journal "Tehnicki Glasnik-Technical Journal" has been running since October 2019 on Publons (<https://publons.com/journal/166059/tehnicki-glasnik-technical-journal>). What is Publons? The following is a text taken from Publons pages or similar ones that best describe what Publons is.

"Publons is a commercial website that provides a free service for academics to track, verify, and showcase their peer review and editorial contributions in academic journals. It was launched in 2012 and by 2018 more than 500,000 researchers have joined the site, adding more than one million reviews across 25,000 journals."

"Publons' mission is to 'speed up science by harnessing the power of peer review.' Publons claims that by turning peer review into a measurable research output, academics can use their review and editorial record as evidence of their standing and influence in their field. Publons says its business model is based on collaborating with publishers."

"Publons produces a verified record of a person's review and editorial activity for journals. This evidence is showcased on reviewers' online profiles and can be downloaded in order to be included in CVs, funding and job applications, and promotion and performance evaluations."

"Publons also provides: tools for publishers to find, screen, contact, and motivate peer reviewers; data and publications about global peer review behaviour; peer review training for early-career researchers; and features for academics to discuss and evaluate published research."

"Reviewers can choose whether to make the content of their reviews openly accessible following the publication of the reviewed publication. However, journals can choose to override this. Review content is shared by using a Creative Commons CC BY 4.0 license. Publons has partnerships with major publishers, including Springer Nature, Taylor and Francis, Oxford University Press, BMJ, SAGE, Wiley and more, and with related services such as Altmetric and ORCID."

"Clarivate Analytics acquired Publons in 2017. The Clarivate's citation network and researcher tools, including the products such as Web of Science, EndNote, and ScholarOne, has generated customers in more than 100 countries and 7,000 institutions. Publons - which continues to run as a stand-alone business - says that partnership will bring more transparency, recognition, and training to peer reviews, helping to solve some of the critical problems that research faces today. This includes speeding up the research process by helping editors find skilled and motivated peer reviewers, and tackling issues such as peer review fraud."

"Please, join over 2,000,000 researchers on Publons in order to track your publications, citation metrics, peer reviews, and journal editing work in a single, easy-to-maintain profile."

"Publons Peer Review Awards are recognitions for top peer reviewers and editors. Publons' Awards began in 2016. In 2017, an award program called the Sentinel Award was added, and it is given for outstanding advocacy, innovation or contribution to scholarly peer reviews."

In this issue, we published 16 papers from various technical fields, six of them were orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC 2019), held in Phuket (Thailand) from October 30 to November 1. These six articles were nominated in the top 10 best articles presented at the conference. Following the usual review process and proofreading, they are accepted for publication in our journal and they will not be published anywhere else.

In terms of classified articles, in this issue we published 11 original scientific papers and 5 preliminary communications.

Best regards,

Milan Kljajin
Editor-in-Chief of **Technical Journal**

SLEEP APNEA DETECTION USING DEEP LEARNING

Hnin Thiri CHAW, Sinchai KAMOLPHIWONG, Krongthong WONGSRITRANG

Abstract: Sleep apnea is the cessation of airflow at least 10 seconds and it is the type of breathing disorder in which breathing stops at the time of sleeping. The proposed model uses type 4 sleep study which focuses more on portability and the reduction of the signals. The main limitations of type 1 full night polysomnography are time consuming and it requires much space for sleep recording such as sleep lab comparing to type 4 sleep studies. The detection of sleep apnea using deep convolutional neural network model based on SPO2 sensor is the valid alternative for efficient polysomnography and it is portable and cost effective. The total number of samples from SPO2 sensors of 50 patients that is used in this study is 190,000. The performance of the overall accuracy of sleep apnea detection is 91.3085% with the loss rate of 2.3 using cross entropy cost function using deep convolutional neural network.

Keywords: continuous single bio-parameter recording; deep convolutional neural network; deep learning; type 4 sleep study; portable sleep apnea detection

1 INTRODUCTION

The type of studies used in our system is type 4 sleep studies, which also refers to continuous single bio-parameter or dual-bio parameter recording. The minimum number of signals that can be used in this type 4 studies is one or two channels such as oxygen saturation and airflow [3-6, 12-30]. Normally type 4 studies do not have EEG and EMG signals, so scoring sleep is not possible. It can be used to identify respiratory sleep disorder during sleep time. The use of oxygen saturation and tracheal sound via acoustic sensor is another alternative to type 4 sleep studies. The primary purpose of using the deep convolutional neural network model for the detection of sleep apnea is to learn parameters or features for the model from the training dataset [7]. The performance of the model is related to the learned parameters in order to determine whether the sample of the signal contains sleep apnea event or not [7-9]. The relationship between the interpretability of the different parameters and the performance of the deep convolutional model is a challenging task for the detection of sleep apnea. A variety of different convolutional neural networks can be used for image classification task; however, only a smaller percentage of them is used for the detection relating to sound waves or the signal processing. The designing and testing of good convolutional neural network model is similar to the black box and needs to be done in a substantial amount of trials and errors because of the inner mechanism of the CNN model [9]. The representation of the signals from the sensor gains a lot of attention in the deep learning model. The proposed model is inspired from the image classification in which the amplitude from the sensors wave data can be seen as the digit vector figures, which can change to any dimensional shape for matrix vectorization or factorization. The main challenge of the sleep apnea detection is that the signal representation is not clear due to continuous nature of the high frequency breathing signal waves. The objective of proposing the biological signal to deep learning model is to fill the gap between the nature of deep learning and continuous nature of the biological signal from the sensors. The use of

convolutional neural network model is a plausible and proper solution to interpret and learn the time series nature of the high frequency SPO2 signal [30-32]. After carrying out five empirical experiments for the validation and testing of sleep apnea detection, the result shows that it can provide meaningful interpretation for the clinical professional. The major contribution of the paper can be summarized as follows:

- A deep learning approach using convolutional neural network is proposed for the detection of sleep apnea in order to explore the segment of the time series data whether apnea occurs or not.
- The accuracy and loss rate SPO2 signal are compared with other approaches of sleep apnea detection using SPO2 signal.

The rest of this paper is organized as follows. The deep learning steps which include data collection, data preparation and segmentation, selection of deep learning model, training, testing, hyper-parameter tuning and prediction, are illustrated in section 2. The system design of the deep learning model can also be seen in Section 2. The performance evaluation results of our system are discussed in Section 3 with five empirical studies using subject validation of 50 patients. The studies contain 190,000 samples of the patients who have sleep apnea. However, the seriousness level of sleep apnea for each patient is different depending on the AHI index and the condition of the patients. Finally, Section 4 concludes and highlights the potential of future research direction for sleep apnea detection.

2 DEEP LEARNING STEPS

The biological neural network of human brain decision making inspires the proposed deep learning model using convolutional neural network [7, 9, 31, 32]. It is based on the single stimulus in which a number of neurons carries messages through electrochemical process, which is required for decision making. In the real world, the identification and prediction of sleep apnea by deep learning method would

facilitate patients with sleep apnea disease [1]. The deep learning model allows to learn from the feature to represent the nature of the data and its patterns. There are two types of trends in deep learning known as supervised learning, in which training data includes both input and desired output, and unsupervised learning, in which training data contains input but not the desired output [31]. Fig. 1 represents the deep learning process in general such as data collection, data preparation and segmentation, deep learning model design such as deep convolutional neural network, training, testing, hyper-parameter tuning and prediction.

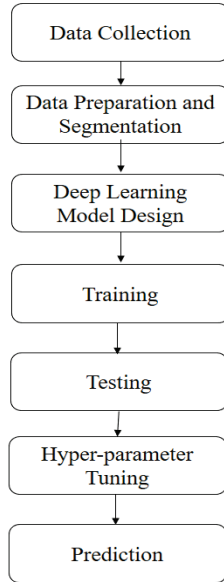


Figure 1 Deep learning steps for apnea detection

2.1 System Design

The design of the detection of sleep apnea can be seen in three layers in which the first layer, second layer and third layer are data collection layer, analysis layer and diagnosis layer respectively and can be seen in Fig. 2.

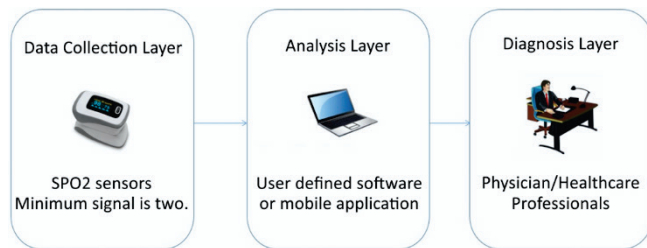


Figure 2 System Design of Sleep Apnea Detection

The first layer is for collecting the data on the oxygen saturation from the patient in order to detect sleep apnea. The second layer is for analysis of the detection of sleep apnea. In the third diagnosis layer, the number of events of sleep apnea occurrence can be identified together with how serious the sleep breathing disorder is for the patients. The third layer is called the diagnosis layer in which health care professional identifies whether the identification of sleep apnea is correct or not.

For one-dimensional feature vector as shown in Fig. 3, the segmented signal will be labelled before being put in the system. For example, if X is assumed as one segmented SPO2 signal, $X_a = \{x_1, x_2, x_3, \dots, x_n\}$ in which $n = \text{sampling rate} * \text{one epoch (30 sec)}$ and $x_n \in X_a$ and $X_a \in X$, where x is a random variable representing the signal data per second. Before putting it into the deep learning model, each segmented signal is stored in the feature vector as follow.

X_1	X_2	X_3	X_{n-1}	X_n	Apnea/ Not Apnea
-------	-------	-------	-------	-----------	-------	------------------

Figure 3 Data Vector of SPO2 Sensor

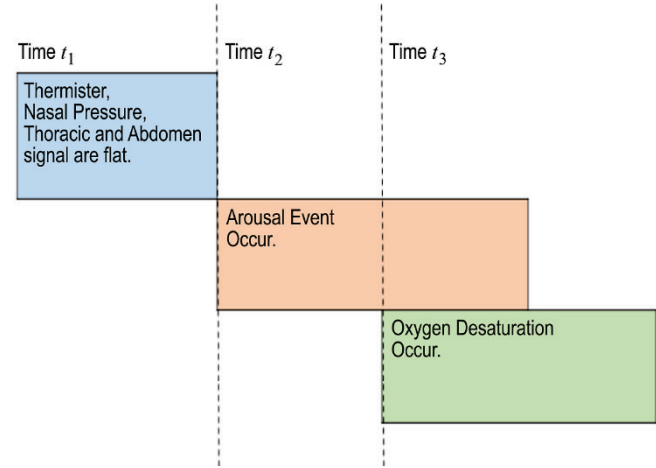


Figure 4 Three consecutive events when sleep apnea occurs

According to Fig. 4, the figure illustrates how the sleep apnea event occur. There are three consecutive time events known as t_1 , t_2 , and t_3 . When breathing stops, the brain sends some signals to the muscle in order to do breathing. This kind of arousal event can be seen in the EEG electrodes' signal that are attached to the patients' head. This event can be seen and is marked in t_2 time segment. The arousal event from the EEG electrode, respiratory signals flat event can be seen in the following Fig. 5 and Fig. 6.

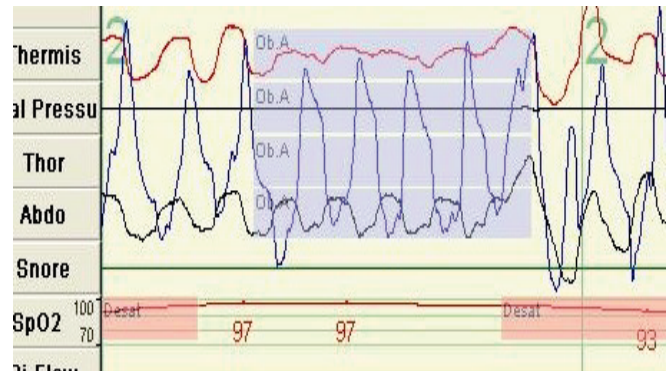


Figure 5 Thermistor, nasal pressure, thoracic and abdominal signals drop event relating to time event

As for the event relating to the time t_1 from Fig. 4, it is significant that the signals coming from the nasal pressure, thermistor, thoracic and abdomen decreased and the signals become flat compared to the normal trend line. This kind of event can be seen in Fig. 5. In other words, these four signals

drop significantly compared to the normal breathing condition. The first four signals highlighted in blue are thermistor, nasal pressure, thoracic and abdominal signal.

After occurrence of the signal drop in these four signals, the oxygen desaturation occurs in SPO2 signal highlighted in red. The abrupt shift of the brain signal from the electrodes with the frequency of less than 16% which does not include spindle and it can last for at least 3 sec. When the sleep apnea occurs, this event is highlighted using green and a red box in Fig. 6.

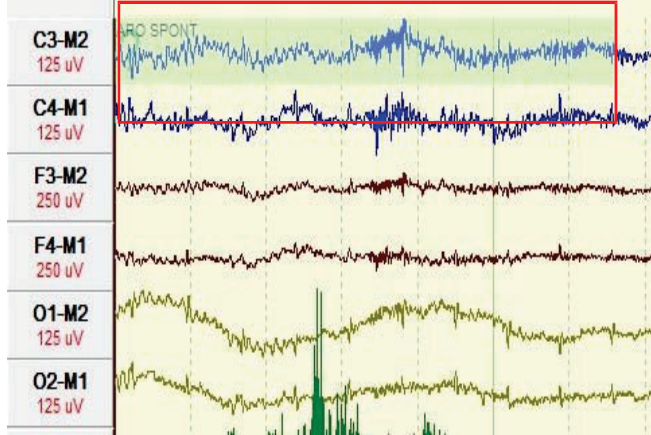


Figure 6 Arousal event relating to time event t_1

As a result of the lung stop breathing, the occurrences of oxygen desaturation can be seen in the t_3 time event of SPO2 signals relating to Fig. 6. The C3-M2 means the central electrode on the left side of the brain is referenced to the right pre-auricular electrode near the right ear. The C4-M1 means the central electrode on the right side of the brain is referenced to the left pre-auricular electrode near the left ear. The F3-M2 means the frontal electrode on the left front side of the brain is referenced to the right pre-auricular electrode near the right ear. The F4-M1 means the frontal electrode on the right side of the brain is referenced to the left pre-auricular electrode near the left ear. The O1-M2 means the occipital electrode on the left front side of the brain is referenced to the right pre-auricular electrode near the right ear. The O2-M1 means the occipital electrode on the right side of the brain is referenced to the left pre-auricular electrode near the left ear. One of the findings is that the heart rate began to increase after the oxygen desaturation event happened in time t_3 .

2.2 Apnea/Hypopnea Event Detection

The flat signals do not mean that the signal is very flat because the up and down can be seen according to minute, second and hour view. The baseline amplitude for the apnea is the preceding 3 to 6 breaths in which time segment before and after the event is represented as t_1 and t_3 respectively, and the amplitude before and after the apnea event is represented as A_1 and A_2 respectively in which $t_1 < t_2 < t_3$. If the maximum relative amplitude for the apnea threshold is set to 10 percent, it means that A_2 divided by A_1 should be less than 10 percent

and the apnea event is detected if the amplitude of the event is twenty percent or below of preceding data according to Fig. 7. For the hypopnea event the maximum threshold is less than 70 percent according to AASM manual scoring 2017.

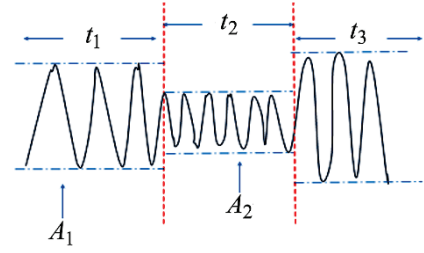


Figure 7 Amplitude Threshold Apnea Event.

The following Fig. 8, illustrates that the minimum duration of the event that can be identified as apnea or hypopnea is ten seconds. It is highlighted in red bounding box in which T is equal to ten seconds.

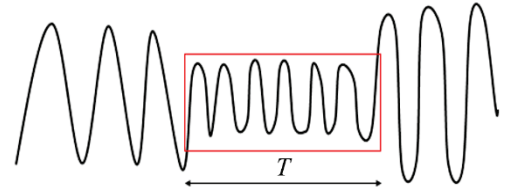


Figure 8 Amplitude Threshold for the Apnea Event

2.3 Deep Convolutional Neural Network

The simplified version of the convolutional neural network is portrayed in Fig. 9 in which the first layer is the input layer, the last layer is the output layer which can be identified whether certain segment from the sensor has apnea or not, and the layers between the first and last layer are the hidden layers.

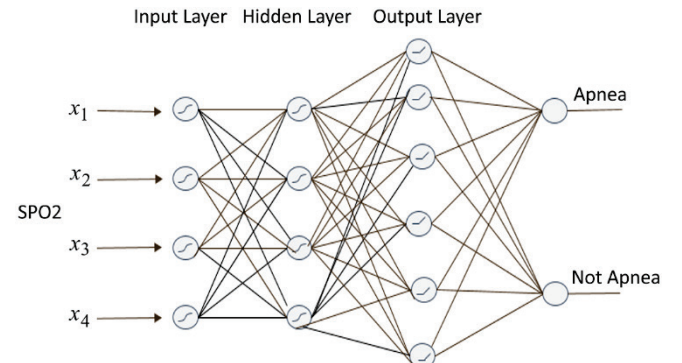


Figure 9 A simple representation of CNN for SPO2

The following Tab. 1 represents the summary of the design of the deep convolutional neural network in which ten layers are used. Moreover, it includes three two-dimensional convolutional layers and three max pooling layers. After running the deep convolutional neural network model for each empirical study, there will be 32,614 trainable parameters for the convolutional neural network and 794 samples are used for training and 3,178 samples are used for

testing. Each sample has 484 data points of SPO2 signal with the frequency of 16 Hz. The proposed model includes pooling and two-dimensional convolutional layers, the flatten layer, dense layer and dropout layer. There are 320 parameters for the first convolutional layer, 9,248 parameters for the second convolutional layer and 18,496 parameters for the third convolutional layer. In addition, there are 4,160 parameters for the first dense layer and 390 parameters for the third dense layer. The nature of the convolutional layer is to extract features, and to learn the special characteristic of the input signal. The pooling layer that takes the maximum value reduces the dimensionality of the model without losing the features or pattern. In addition, the fully connected layer is also known as the dense layer, which can be used as a final layer for prediction.

Table 1 Deep Convolutional Neural Network Design

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 20, 20, 32)	320
max_pooling2d_1	(MaxPooling2D(None,10,10,32))	0
conv2d_2 (Conv2D)	(None, 8, 8, 32)	9248
max_pooling2d_2	(MaxPooling2D(None,4,4,32))	0
conv2d_3 (Conv2D)	(None, 2, 2, 64)	18496
max_pooling2d_3	(MaxPooling2D(None, 1, 1, 64))	0
flatten_1 (Flatten)	(None, 64)	0
dense_1 (Dense)	(None, 64)	4160
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 2)	390
Total params: 32,614		
Trainable params: 32,614		
Non-trainable params: 0		
Train on 794 samples, validate on 3178 samples		

3 PERFORMANCE EVALUATION

It can be seen that five empirical studies are made using the accuracy metrics of the proposed deep convolutional neural network model using fifty patients and each empirical study includes 3,700 to 4,000 sample data of ten patients. Each empirical study is illustrated in two figures: one for accuracy and the other for loss. The cross entropy is used for calculating the loss rate. The overall performance of the proposed model is 91.3085 percent. The comparison between proposed convolutional neural network model and the other models can also be found in this performance evaluation section. It contains the comparison different classifier such as linear discriminant analysis (LDA), support vector machine (SVM), bagging representation tree, artificial neural network using SPO2 signal and the proposed convolutional neural network of deep learning model.

The summary of the comparison of five empirical studies regarding training accuracy can be seen in Fig. 10 in which scenario 1, scenario 2, scenario 3, scenario 4 and scenario 5 are illustrated using orange, red, green, purple and blue trend line respectively. Each scenario includes ten patients with difference age, BMI, weight, AHI index. All of the trend lines fluctuated and gradually increased between 1 percent and 20 iterations. The percentage of accuracy starts to stabilize after twenty iterations; they become more stable after 40 iterations with the accuracy around 90 percent. It can be concluded that

the stable condition for training the data is between 40 iterations to 100 iterations.

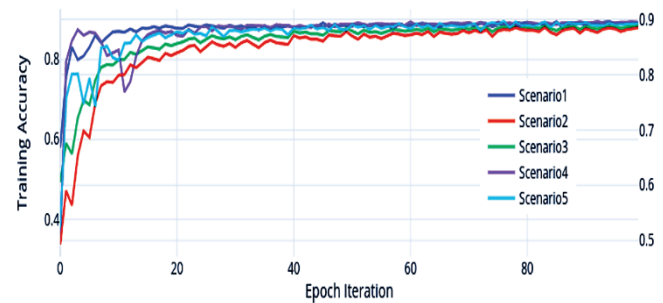


Figure 10 Comparison of training accuracy for five empirical studies

The summary of the comparison of five empirical studies regarding training prediction loss can be seen in Fig. 11 in which scenario 1, scenario 2, scenario 3, scenario 4 and scenario 5 are illustrated using blue, orange, green, purple and cyan trend line respectively. In this paper, the total number of patients that participated in the detection of sleep apnea is fifty patients and subject specific scenario validation is used with the split rate of 0.2. Each of the scenarios had ten different patients of different age, BMI, weight, AHI index. All of the trend lines relating to training loss were going down until the number of twenty iterations and at the same time, they fluctuated and were unstable until twenty iterations. They became more stable after 20 iterations. The stable condition for training prediction loss was between 20 iterations to 100 iterations.

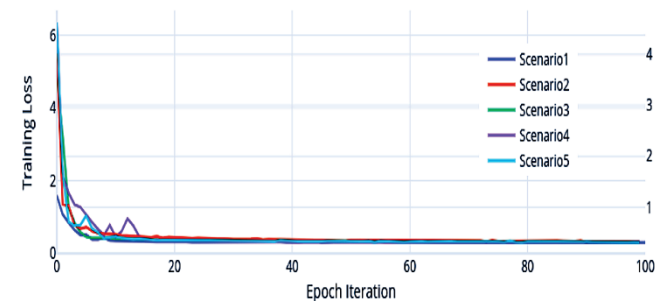


Figure 11 Comparison of training prediction loss for five empirical studies

Table 2 Comparison between different types of classifier for sleep apnea detection.

Ref	Type of Classifier	Split Rate	Training Data	Testing Data	Accuracy
[29]	LDA	0.5	50%	50%	86.5
[30]	SVM	0.3	70%	30%	90
[31]	Bagging Rep Tree	0.1	90%	10%	84.80
[32]	Artificial Neural Network	0.17	83%	17%	90.3
Proposed Model	CNN	0.2	20%	80%	91.3085

According to Tab. 2, the SPO2 signals are used for the comparison between the proposed model and other classifiers. The proposed model outperformed the other models with the accuracy of 91.3085 percent with the split

rate of 0.2 in which the training data is 20 percent (794 samples) and testing data is 80 percent (3178 data samples).

The following Tab. 2 illustrates the comparison between different types of classifier such as LDA, SVM, Bagging representation tree, artificial neural network for sleep apnea detection. It can be seen that deep convolutional neural network has the accuracy of 91.3085 percent. In other word, the model can predict very well for 913085 data out of 100,0000 data. The other model use only a few percentage data for testing and a huge percentage of data for training. However, the twenty percent of data are used for the training and 80 percent of data are used for testing in order to avoid overfitting in the proposed model for detection of sleep apnea. Actually, the twenty percent used for training is not small enough for the proposed model because of 8 hours of sleep recording for each patient. Fifty patients involve in this five empirical studies. The other model uses the standard database with a few amount of data, however proposed model use own database after collecting data of fifty patients who have sleep apnea with the help of health care professionals from the sleep lab.

4 CONCLUSION

As a summary, the preparation and segmentation the data model is a crucial step before feeding into the deep learning. The experiment of deep learning using deep CNN (convolutional neural network) model is tested using SPO2 using subject validation with the split rate of 0.2 in order to avoiding overfitting problem. The overall accuracy for the entire sample using subject validation is 91.3085 in which the trend line for testing is better than the trend line for training

Acknowledgement

This work was supported by Thailand's Education Hub for Southern Region of ASEAN Countries (TEH-AC), Prince of Songkla University Graduate Studies Grant as well as by Department of Computer Engineering, Prince of Songkla University. The authors would like to express the gratitude for our supervisor, co-supervisor and sleep technicians from Songklanagarind Hospital, Thailand for guiding in my research work and supporting access to the respiratory signals acquired in Sleep Disorders Service and Research Center.

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

5 REFERENCES

- [1] Huang, W., Guo, B., Shen, Y., & Tang, X. (2017). A novel method to precisely detect apnea and hypopnea events by airflow and oximetry signals. *Computers in Biology and Medicine*, 88(June), 32-40. <https://doi.org/10.1016/j.combiomed.2017.06.015>
- [2] Berry, R. B., Gamaldo, C. E., Harding, S. M., Brooks, R., Lloyd, R. M., Vaughn, B. V., & Marcus, C. L. (2015). AASM scoring manual version 2.2 updates: New chapters for scoring infant sleep staging and home sleep apnea testing. *Journal of Clinical Sleep Medicine*, 11(11), 1253-1254. <https://doi.org/10.5664/jcsm.5176>
- [3] Ravesloot, M. J. L., Frank, M. H., van Maanen, J. P., Verhagen, E. A., de Lange, J., & de Vries, N. (2016). Positional OSA part 2: retrospective cohort analysis with a new classification system (APOC). *Sleep and Breathing*, 20(2), 881-888. <https://doi.org/10.1007/s11325-015-1206-y>
- [4] Lazareck, L. & Tarassenko, L. (2006). Detection of apnoeic and breathing activity through pole-zero analysis of the SpO2 signal. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, 6573-6576. <https://doi.org/10.1109/IEMBS.2006.259623>
- [5] Vázquez, J.-C., Tsai, W. H., Flemons, W. W., Masuda, A., Brant, R., Hajduk, E., Whitelaw, W. A., & Remmers, J. E. (2000). Automated analysis of digital oximetry in the diagnosis of obstructive sleep apnoea, 302-307. <https://doi.org/10.1136/thorax.55.4.302>
- [6] DeMeulenaere, S. (2007). Pulse Oximetry: Uses and Limitations. *The Journal for Nurse Practitioners*, 3(5), 312-317. <https://doi.org/10.1016/j.nurpra.2007.02.021>
- [7] Zeng, H., Haleem, H., Plantaz, X., Cao, N., & Qu, H. (2017). CNN Comparator: Comparative Analytics of Convolutional Neural Networks.
- [8] Almazaydeh, L., Elleithy, K., Faezipour, M., & Abushakra, A. (2013). Apnea Detection Based on Respiratory Signal Classification. *Procedia - Procedia Computer Science*, 21, 310-316. <https://doi.org/10.1016/j.procs.2013.09.041>
- [9] Ng, A. S. L., Chung, W. Y. J., Di Gohel, M., Fan, K. L., & Wong, T. K. S. (2008). Evaluation of the performance of using mean absolute amplitude analysis of thoracic and abdominal signals for immediate indication of sleep apnoea events. *Journal of Clinical Nursing*, 17(17), 2360-2366. <https://doi.org/10.1111/j.1365-2702.2008.02323.x>
- [10] Ng, E. K., Koh, T. S., Baey, E., Lee, T. H., Abeyratne, U. R., & Puvanendran, K. (2008). Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea? *Sleep Medicine*, 9(8), 894-898. <https://doi.org/10.1016/j.sleep.2007.07.010>
- [11] Ng, A. K., Koh, T. S., Baey, E., & Puvanendran, K. (2006). Speech-like analysis of snore signals for the detection of obstructive sleep apnea. *ICBPE 2006 - Proceedings of the 2006 International Conference on Biomedical and Pharmaceutical Engineering*, 99-103.
- [12] Lazareck, L. & Tarassenko, L. (2006). Detection of apnoeic and breathing activity through pole-zero analysis of the SpO2 signal. *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 6573-6576. <https://doi.org/10.1109/IEMBS.2006.259623>
- [13] Huang, W., Guo, B., Shen, Y., & Tang, X. (2017). A novel method to precisely detect apnea and hypopnea events by airflow and oximetry signals. *Computers in Biology and Medicine*, 88(June), 32-40. <https://doi.org/10.1016/j.combiomed.2017.06.015>
- [14] Mostafa, S. S., Mendonça, F., Morgado-Dias, F., & Ravelo-García, A. (2017). SpO2 based Sleep Apnea Detection using Deep Learning. *21st International Conference on Intelligent Engineering Systems*, 91-96.
- [15] Zhang, J., Zhang, Q., Wang, Y., & Qiu, C. (2013). A Real-Time Auto-Adjustable Smart Pillow System for Sleep Apnea Detection and Treatment. In *Proceedings of the 12th*

- International Conference on Information Processing in Sensor Networks (IPSN)*, 179-190.
<https://doi.org/10.1145/2461381.2461405>
- [16] Lin, Y.-Y., Wu, H.-T., Hsu, C.-A., Huang P.-C., Huang, Y.-H., & Lo, Y.-L. (2017). Sleep Apnea Detection Based on Thoracic and Abdominal Movement Signals of Wearable Piezoelectric Bands. *IEEE Journal of Biomedical and Health Informatics*, 21(6), 1533-1545. <https://doi.org/10.1109/JBHI.2016.2636778>
- [17] Camci, B., Kahveci, A. Y., Arnrich, B. & Ersoy, C. (2017) Sleep apnea detection via smart phones. *25th Signal Processing and Communications Applications Conference*.
<https://doi.org/10.1109/SIU.2017.7960484>
- [18] Mostafa, S. S., Carvalho, J. P., Morgado-Dias, F., & Ravelo-García, A. G. (2017), Optimization of Sleep Apnea Detection using SpO2 and ANN. *26th International Conference on Information, Communication and Automation Technologies*.
<https://doi.org/10.1109/ICAT.2017.8171609>
- [19] Jayaraj, R., Mohan, J., & Kanagasabai, A. (2017). A Review on Detection and Treatment Methods of Sleep Apnea. *Journal of Clinical and Diagnostic Research: JCDR*, 11(3), 1-3.
<https://doi.org/10.7860/JCDR/2017/24129.9535>
- [20] Du, N. & Liu, K. (2017). ApneaRadar : A 24 GHz Radar-Based Contactless Sleep Apnea Detection System. *2nd International Conference on Frontiers of Sensors Technologies (ICFST)*, 2-6. <https://doi.org/10.1109/ICFST.2017.8210539>
- [21] Rolón, R. E., Larrateguy, L. D., Di Persia, L. E., Spies, R. D., & Rufiner, H. L. (2017). Discriminative methods based on sparse representations of pulse oximetry signals for sleep apnea-hypopnea detection. *Biomedical Signal Processing and Control*, 33, 358-367. <https://doi.org/10.1016/j.bspc.2016.12.013>
- [22] Sharma, H. & Sharma, K. K. (2016). An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions. *Computers in Biology and Medicine*, 77, 116-124.
<https://doi.org/10.1016/j.compbiomed.2016.08.012>
- [23] Ma, H. T., Ieee, M., Liu, J., Zhang, P., Zhang, X., & Yang, M. (2015). A Real-Time Automatic Monitoring System for Sleep Apnea Using Single-Lead Electrocardiogram. *IEEE Transactions on Information Technology in Biomedicine*, 1-4.
<https://doi.org/10.1109/TENCON.2015.7372966>
- [24] Shiota, S., Inoue, Y., Takekawa, H., Kotajima, M., Nakajyo, M., Usui, C., & Takahashi, K. (2014). Effect of continuous positive airway pressure on regional cerebral blood flow during wakefulness in obstructive sleep apnea. *Sleep & Breathing*, 18(2), 289-295. <https://doi.org/10.1007/s11325-013-0881-9>
- [25] Arzt, M., Woehrle, H., Oldenburg, O., Graml, A., Sulin, A., Erdmann, E., & Wegscheider, K. (2016). Prevalence and Predictors of Sleep-Disordered Breathing in Patients with Stable Chronic Heart Failure. *JACC: Heart Failure*, 4(2), 116-125. <https://doi.org/10.1016/j.jchf.2015.09.014>
- [26] Tu, J. F. (2017). A breathe regulator used to improve the breathe obstacle when sleeping. *Proceedings of the IEEE International Conference on Advanced Materials for Science and Engineering: Innovation, Science and Engineering, IEEE-ICAMSE 2016*, 649-652.
<https://doi.org/10.1109/ICAMSE.2016.7840272>
- [27] Pierleoni, P., Pernini, L., Belli, A., & Palma, L. (2014). Real-time apnea detection using pressure sensor and tri-axial accelerometer. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 513-516.
<https://doi.org/10.1109/BHI.2014.6864415>
- [28] Lee, H. K., Lee, J., Kim, H., & Lee, K. J. (2013). Automatic snoring detection from nasal pressure data. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6870-6872.
- [29] Ravelo-García, A. G., Kraemer, J. F., Navarro-Mesa, J. L., Hernández-Pérez, E., Navarro-Esteva, J., Juliá-Serdá, G., & Wessel, N. (2015). Oxygen saturation and RR intervals feature selection for sleep apnea detection. *Entropy*, 17(5), 2932-2957.
<https://doi.org/10.3390/e17052932>
- [30] Zhang, J., Zhang, Q., Wang, Y., & Qiu, C. (2013). A Real-Time Auto-Adjustable Smart Pillow System for Sleep Apnea Detection and Treatment. In *Proceedings of the 12th International Conference on Information Processing in Sensor Networks (IPSN)*, 179-190.
<https://doi.org/10.1145/2461381.2461405>
- [31] Xie, B. & Minn, H. (2012). Real-Time Sleep Apnea Detection by Classifier Combination. *IEEE Trans. Inf. Technol. Biomed.*, 16(3), 469-477. <https://doi.org/10.1109/TITB.2012.2188299>
- [32] Elleithi, K., Almazaydeh, L., & Faezipour, M. (2012). A Neural Network System for Detection of Obstructive Sleep Apnea through SpO2 Signal Features. *Int. J. Adv. Comput. Sci. Appl.*, 3(5), 7-11. <https://doi.org/10.14569/IJACSA.2012.030502>

Authors' contacts:

Hnin Thiri CHAW
 Dept of Computer Engineering,
 Prince of Songkla University,
 90112 Hat Yai, Thailand
hninthirichaw87@gmail.com

Sinchai KAMOLPHIWONG
 Dept of Computer Engineering,
 Prince of Songkla University,
 90112 Hat Yai, Thailand
ksinchai@psu.ac.th

Krongthong WONGSRITRANG
 Dept. of Otolaryngology Head Neck Surgery,
 Songklanagrin Hospital,
 90112 Hat Yai, Thailand
golf@hotmail.com

THE EMPIRICAL STUDY: ENCOURAGING STUDENTS' INTEREST IN SOFTWARE DEVELOPMENT USING TEST-DRIVEN DEVELOPMENT

Aziz NANTHAAMORNPONG, Stephane BRESSAN

Abstract: The supply is not matching the demand on the market for software developers. While the enrolment in undergraduate computer science courses is increasing, few students are interested in and committed to becoming software developers. It could be that students are overwhelmed by the software development methodology that they are taught. We are consequently looking for a constructivist approach to software engineering able to effectively engage learners. We empirically evaluated whether test-driven development (TDD) is able to improve the quality of both learning and of software development in the classroom. Although numerous studies have outlined the benefits and effects of TDD in the classroom, none of those studies have focused on measuring students' interest in and attitudes toward using TDD in the classroom. We present a study evaluating the impact of TDD on the engagement and focus of learners of software development in the classroom. The results illustrate that the use of TDD in the classroom encourages learners to engage and focus.

Keywords: empirical software engineering; software engineering education; test-driven development

1 INTRODUCTION

"By 2020, one of every two jobs in science, technology, engineering, and mathematics will be in computing," according to Bobby Schnabel, executive director and chief executive officer of the Association for Computing Machinery [1]. The enrolment in undergraduate computer science courses is increasing. Yet, the demand, the number of jobs in computing, outruns the supply, the number of students graduating in computer science and related fields, according to the report "Assessing and responding to the Growth of Computer Science Undergraduate Enrolments" by the American National Academies of Science, Engineering and Medicine [2]. Stack Overflow 2019 Developer Survey Results [3] suggest that the majority of software developers are Web developers (although this may be a bias inherent to the Stack Overflow community) and that they are not necessarily computing graduates. This highlights the existence of a market for trained software developers able to work on challenging applications such as distributed and embedded systems.

If well managed, this phenomenon is an opportunity for all countries with the right education policies and the right academic curricula and the right methodologies. Developing countries, in particular, can diversify their industry and workforce and further tune into the growing service and knowledge economy. Illustratively, Oichi Okoshi, president and chief executive officer of Toyota Tsusho Nexty Electronics (Thailand) Co., Ltd. believes that "Thailand has a lot of potential to become a hub of automotive software development" [4].

This is however even more challenging than it looks as, while the number of students wanting to study computing and pursue a career in computing is increasing, it is not necessarily the case that a sufficient number of these students are interested in coding and are actually motivated and able to effectively learn to code. In its report [2], the American National Academies of Science, Engineering and Medicine

further acknowledges that academically trained software developers end up not being properly equipped to meet the requirements of the software industry.

Typically, the teaching pattern of software development starts with programming fundamentals such as variable declaration and assignment, control structures (e.g., if-then-else, for and while statements) and other language structures (e.g., functions), after which exercises or problems are assigned to the students. This pattern parallels a well-known software development process: The Waterfall Model [5]. In this model, software development begins by acquiring Software Requirements for Software Design. Implementation then starts with code writing. Finally, Testing is performed once the code is complete. Both the teaching methodology and the Waterfall model emphasize a conceptually top-down approach from the abstract to the concrete. Consequently, students focus on executing software with no runtime error. Diving from the height of the software requirement into the code without anything to guide them, students can only rely on a trial and error approach [6] in which debugging is the main activity. They actually generally test their programs with only a handful of test cases after the code is finished. Testing of all operational conditions is rarely performed. Paradoxically, testing is assumed to be completed when the program runs without error. As a result, software development projects fail in practice. In addition, the time spent on software development is increasing because fixing bugs is time consuming, albeit the main activity.

In this top-down approach, students misunderstand system requirements or cannot find a solution to the assigned global problem. They cannot write code to meet the requirements and get hopelessly lost in meaningless debugging of program errors and random attempts to fix failures to run.

Piaget's theory of cognitive development [7] was the seminal impulse to constructivist pedagogy [8]. Constructivist instructional approaches try and actively

involve learners in the construction process of meaning and knowledge. Clearly, constructivism suggests bottom-up approaches. We look among the modern agile software engineering methods one that can provide both the active involvement and the rigor needed.

Test-Driven Development (TDD) is a method of software development and is part of Extreme Programming, which is an agile software development method [9]. Agile software development has been a popular method in the global software development community since 2001. Because of the complexity of software systems, the popularity of agile development has been increasing. Agile development can synchronize communications among developers, users and customers regarding software requirements. Essentially, agile software development operates in a short cycle, and the developer selects numerous functionalities to develop in each cycle. Once the chosen functionality has been developed and delivered to the customers, those customers can rapidly perceive the tangible outcomes that satisfy their requirements. The customers can give feedback and suggestions from the beginning of the development process, significantly decreasing the risk that the system will not satisfy their requirements. The result is higher software quality [10]. Additionally, TDD is recommended for teaching students about software development because TDD helps improve the software development process [11], [12]. The researchers seem to agree that TDD is an enjoyable way to work on the programming task [13]. The TDD approach involves writing the test code before the functional code. Therefore, the developers write the code for testing before writing the actual functional code and test the production code using the created testing code. Programming is done in cycles for every function. After passing the test, the developers can improve the quality of the code, since the work in this stage is only to ensure that the code works properly; the result can be a poor-quality program. Updating the code is called Refactoring. The refactoring process changes the internal structure of the software but does not alter the functionality and is done to enhance the quality of the software – for example, by addressing feature envy – which can be refactored using ‘move method’ [14, 15]. Repetitive tasks can be implemented as a superclass; any function can then call this super-class from the regular class, helping prevent duplication and reduce lines of code.

As far as we are aware, there is a lack of empirical evidence presented in the literature as to whether TDD can be instrumental in sparking student, particularly undergraduate student, interest in software development. Therefore, we are interested in TDD adoption by students in software development studies. The main research question of this study is as follows: *Can test-driven development encourage student interest in software development?*

2 RELATED WORK

Janzen suggested that academics benefit from adopting the TDD teaching approach to teach testing. Textbooks and instructional materials are expected to incorporate this

approach, since TDD can help improve software development [16]. In later years, Janzen and Saiedia [17] investigated whether the use of TDD in teaching and learning can lead to better software systems. The study indicates that although software quality is better, the research still lacks an examination of students’ feelings and attitudes about the use of TDD for assigned tasks. Another study by Pencur et al. [18] involved two groups of developers. Group 1, Iterative test-last (ITL), used the traditional testing method, which is to test the program after the code was finished. Group 2 incorporated TDD. The comparison shows that TDD is not substantially different from ITL on code coverage but is significantly different in quality in that the use of TDD results in better code quality.

Edwards [19] evaluated the effects of using TDD when teaching undergraduate students in the classroom and found that students demonstrated the effectiveness and accuracy of the test. The students obtained higher scores and encountered fewer programming errors. Another study was conducted by Buffardi with the goal of motivating new programmers to use TDD [20]. The result indicates that TDD helps produce higher-quality code.

The study of TDD use among professional developers shows that it is a well-respected tool in the software industry [21]. Preliminary reports on TDD testing with experts in software development illustrate that TDD helps improve software quality and encourages rigorous testing. This evidence suggests that the trend among developers in the use of TDD enhances testing capabilities in the software industry [22], [23].

A comparative study of TDD use among university students and expert groups that utilized TDD in real-world environments showed that compared to the student group, the expert group paid more attention to and preferred to use TDD [24].

Desai et al. [11] suggested that TDD is a possible solution for improving students’ software testing skills. They also showed that TDD exposes students to analytical and comprehension skills, both of which are needed in software testing. Mäkinen and Münch [25] performed a comparative analysis of empirical studies on the impact and potential of using TDD through observation. The results from the code quality comparison suggested that TDD can help reduce errors and improve both quality and maintenance time.

Causevi et al. [26] reviewed the literature on the limitations of adopting TDD in the software industry. Based on empirical evidence from 48 studies, the researchers identified the following factors as limiting the industrial adoption of TDD: 1) Increased development time; 2) Insufficient TDD experience/knowledge; 3) Insufficient design; 4) Insufficient developer testing skills; 5) Insufficient adherence to TDD protocol; 6) Domain- and tool-specific limitations; and Legacy code.

3 RESEARCH METHODOLOGY

To address the research question described in Section 1, we identify the following hypotheses: H_{null} : the student’s interest with the TDD method is not different from student

interest with the waterfall model; $H_{alternative}$: the student's interest with the TDD method is different from student interest with the waterfall model.

In this study, we primarily conducted controlled experiments and interviews. The following subsections will detail the experimental setting and data-collection method.

3.1 Experimental Variables

Based on our hypotheses, we have set an independent variable as the *software development* method, including 1) the waterfall model and 2) the TDD method. The dependent variable is *student's interest* in software development.

The fields of psychology and education students have very different definitions of "interest." In this study, we defined interest as follows: *Interest is a feeling of satisfaction or attitude toward anything, any idea or any situation with a tendency of wanting to approach and know about that thing. This leads to perseverance to achieve the goal. Interest is the driving force that motivates people to do anything or represents the tendency for the person to choose. In terms of academics, interest is linked to academic success and is a key element in career skill development* [27], [28]. We used this definition to build the interest indicators for measuring the dependent variable. We have identified 5 indicators of interest as follows:

- 1) Enthusiasm about what is interesting: The purpose of this metric is to identify what students learned before, during, and after the experiment. Being enthusiastic about what interests the students is a driving force that motivates them to choose what they want to study. For example, if students are interested in programming, their interest might motivate them to pursue a career in software development.
- 2) Self-study: This metric provides an overview of what students learned more by themselves, which might be related to software development. It might also be related to software development but is not in the curriculum. For example, students can learn more about Python, which is within the scope of software development but is not taught in any courses at the university.
- 3) Exercises: This indicator studies information from student exercises. The topics for the exercises may be those in which the students are interested. For example, some students might choose to do the system analysis exercise but not the programming exercise.
- 4) Exercise duration: The purpose of this indicator is to study the time spent on the exercises because the amount of time spent by each student is different. For example, a student who takes more time to do exercises or does more exercises than others might want to understand the course material.
- 5) Self-study duration: This metric is intended to study the time used for self-study. For example, students who are interested in learning more about software development in Java may spend more time studying this topic by themselves than other subjects.

Once the interest indicator has been identified, we

adopted metrics to design the questionnaire, which consists of open-ended questions and Likert-scale questions.

3.2 Data Collection

The data were collected using two methods: 1) questionnaires and 2) interviews. The questionnaires and interview questions include both quantitative and qualitative data collection. The data from the students before and after the experiment were analysed both to determine whether the use of TDD in software development can encourage student interest and to study the impact of TDD on education.

Each questionnaire consists of both Likert-scale questions and open-ended questions. The interview questions are based on the open-ended questionnaire because we required in-depth information from the students that might not have been present in their questionnaire answers. This strategy was used because the interviewees are free to answer questions, and answering in their own words is easier than writing formal responses; additionally, they can offer more comprehensive information. Details about the design of the questionnaires and interviews questions are described in the paragraphs to follow.

Questionnaires We developed the questionnaire to provide pre-test questions, post-test questions and TDD exercises. The questions in the questionnaire were examined by a software engineering expert with experience in software development and TDD research. There are two formats of the questions: 1) The open-ended question collects the student's basic information and software development experience in object-oriented programming (OOP) courses and Java projects developed using the waterfall model. It also gathers information about the students' interests in the field of software development, and 2) The Likert-scale question collects information about students' interests in software development. There are five levels of answers: 1) Not at all interested, 2) Slightly interested, 3) Somewhat interested, 4) Very interested, and 5) Extremely interested.

This questionnaire was developed from our defined indicators of interest. Tab. 1-3 show the developed questions with indicators.

Interviews The interview was designed to be semi-structured. In a semi-structured interview, the interviewer can append or alter the questions during the interview. The interview questions are similar to the open-ended questionnaires but are intended to collect additional information from the students' pre-test and post-test. These interviews collect more comprehensive information than the questionnaire since speaking is much easier than writing. In addition, the interview questions are more flexible; thus, the interviewer and interviewee can communicate better, especially when using technical terms.

3.3 PARTICIPANTS

The samples in this study consist of 52 students in the second-year of a software engineering program. We selected

these students because this study involves testing and refactoring, which is part of the students' prior software engineering classes. Additionally, the second-year students have been taught basic programming and OOP. As a result, the students could understand the questions from interviews and questionnaires.

3.4 Controlled Experiment

The controlled experiment consists of four stages, including the pre-test, TDD training, experiment treatments, and post-test.

Pre-test The pre-experiment was performed using questionnaires and interviews. The questionnaire survey was on paper, and the students were asked to answer the survey questions in the classroom beforehand. The students had 30 minutes to complete the questionnaire, which consists of 16 open-ended and closed-ended questions.

The interview questioning was in a semi-structured format in which the interviews were planned in advance. The interviewers included four researchers who were trained by the first author. The content of the interview concerned the students' interest in software development and the software development model used before attending the TDD presentation. If necessary, each interviewer could ask questions in addition to the specified questions. The interview was conducted in a one-to-one format. The interviews were recorded via audio recording and note taking.

Prior to performing the experiment, we conducted the pilot study using designed questionnaires and interviews with 32 students who were not included in the main experiment. The students spent an average of 15 minutes to complete the questionnaire and 20 minutes to be interviewed. The pilot study aimed to examine the internal consistency reliability of the Likert scale questionnaire. We checked the reliability by calculating the Cronbach's alpha coefficient based on the 32 respondents. The calculated coefficient was 0.89, which shows that the magnitude of the consistency of the questions was in the exact same direction (greater than 0.7). Based on this pilot study, we did not find any problems stemming from the questionnaire and interview.

TDD training One week after completing the questionnaire and interview, the second author taught TDD to the students. The topics covered in the lecture were as follows: 1) basic knowledge of TDD, 2) TDD and unit testing, 3) TDD software development tools, and 4) TDD software development examples. All 52 students participated in the lecture session.

Experiment treatments The students were assigned to develop the FizzBuzz program [29], which is an easy program that enabled students to understand the steps in the TDD process. FizzBuzz is a game that helps one learn division by assigning one integer value in four cases as follows:

- **Case 1** for an integer that can divided by 3, display the word Fizz;
- **Case 2** for an integer that can divided by 5, display the word Buzz;

- **Case 3** for an integer that can divided by 3 and 5, display the word FizzBuzz;
- **Case 4** for an integer that cannot be divided by either of the three cases above, display the input integer.

We asked each participant to implement the FizzBuzz game as two versions. For the first version, each participant developed the program using the waterfall model (non-TDD version). In the second version, the program was developed with TDD process (TDD version).

To implement the non-TDD version, the participants could choose any programming software development tools (e.g., NetBean, Eclipse, IntelliJ). For the TDD version, the participants were required to develop a FizzBuzz program using the TDD method with a JUnit plugin (version 4.0) integrated with Eclipse. The participants were asked to install JUnit within Eclipse on their laboratory computers by themselves and test whether JUnit worked properly. Once JUnit was successfully installed, the participants began implementing the program with the TDD process.

The participants developed the non-TDD and TDD versions in the morning (9:00 AM-11:00 AM) and afternoon (1:00-3:00 PM), respectively.

Post-test One day after the assigned tasks were completed, we conducted a survey using questionnaires and interviews. The post-test questionnaire was similar to the pre-test questionnaire, but there were additional questions about the effect of using TDD. The participants were asked to complete the questionnaires within 30 minutes.

4 RESULTS AND DISCUSSION

To examine how well the Likert-scale questions represented the research construct, we employed a statistical method known as confirmatory factor analysis (CFA [30]). The 8 Likert-question scores from the pre- and post-experiments were analysed with the CFA.

First, we measured how our data are suited for a factor analysis. To do so, we used Kaiser-Mayer-Olkin (*KMO*) and Bartlett's test. Typically, if the value of *KMO* < 0.5, it is considered unsuitable to use a factor analysis. Bartlett's test is a statistical hypothesis test. In this study, we assume Bartlett's test in the questionnaire as follows:

H_{null} : The questions are not related

$H_{alternative}$: The questions are related

Based on an analysis of the Likert-scale questionnaire, the *KMO* result equals 0.858, which is greater than 0.5; thus, the data are suitable for use with factor analysis techniques.

Additionally, the value of Bartlett's test is less than 0.05, so H_{null} is rejected and $H_{alternative}$ is accepted. As a result, it can be concluded that the indicators for the questions are correlated, and therefore, the factor analysis can be used. Next, we used principal component analysis (PCA) to examine whether all the questions included all measures of interest of the students. With the recommendation of Kaiser [31], components with eigenvalues greater than 1 should be

retained. In this case, we retain only component 1 as the principal factor.

Table 1 Pre-test questions

Questions	Indicators
1. Grade in OOP subject	Enthusiasm about topics of interest
2. What is your project in OOP about?	
3. What is your reason for enrolling in software engineering?	
4. What do you know about "software developers"?	
5. Are you interested in software development? Why? (Interest is a feeling of satisfaction, curiosity, pursuit of enjoyment, personal satisfaction toward an activity that leads to perseverance to achieve the goal)	
6. What is your language of interest?	
7. How did you test the system in your previous class project?	
8. Based on your past studies, which subjects are significant in the study of software engineering? (Significant subjects can affect your later education or may be required in your software engineering career)	
9. Which subjects have interested you in your previous software engineering classes? (Interest is the satisfaction of learning a subject that you would like to learn more about or study by yourself)	
10. If training is provided as a seminar or workshop, which topic would like to study? Why?	
11. What is your expected career after graduation?	
12. What is the average amount of time you spend on self-study? (average time per week)	
13. What is your preferred subject when you study by yourself?	
14. If training is provided as a seminar or workshop, which topic would you like to study? Why?	
15. Which kind of an assigned exercise?	Exercises
16. What is the average amount of time you spend to develop software from each assigned exercise? (average time per week)	Exercise duration

Table 2 Likert-Scale Questions

Questions	Indicators
1. You are interested in learning software development.	Enthusiasm about topics of interest
2. You are interested in practicing software development from the exercises (Software development practice is in the form of a coding exercise).	
3. You are interested in practicing software development via self-study (Software development practice is in the form of a coding exercise).	
4. You are interested in learning about software engineering by yourself (Studying software engineering topics such as learning a new method of software development or a new programming language, etc.).	
5. You are interested in modifying the interface from the assigned exercise.	
6. You are interested in participating in questions and answers during class.	
7. You are interested in self-learning.	Self-study and self-study duration
8. You are interested in spending time to complete the exercise.	Exercises and exercise duration

We then calculated a factor loading that illustrates the relationship between variables (questions) and the principal

component (component 1). The loading factor values of all variables are greater than 0.5; thus, all variables have relationships within the principal component. Additionally, the Cronbach's alpha values of all questions are 0.929, suggesting that all five indicators were within the same factor and that the variables were highly correlated.

Table 3 Post-Test Questions

Questions	Indicators
1. Does the use of TDD affect your interest in software development? Why?	Enthusiasm about topics of interest
2. How does the use of TDD affect your programming experience in the laboratory?	
3. When using the waterfall model or TDD in the same exercise, which method produces fewer errors? Why?	
4. In the future, if you are required to create a project in class, which test method will you use?	
5. Which method do you believe provides more details in program testing? Why?	
6. From your perspective, does TDD contribute to software development? Why?	

4.1 Likert-Scale Questionnaire

Based on the 52 responses, the average score of each question was summarized by comparison of the mean scores of the pre- and post-experiments, which is shown in Fig. 1.

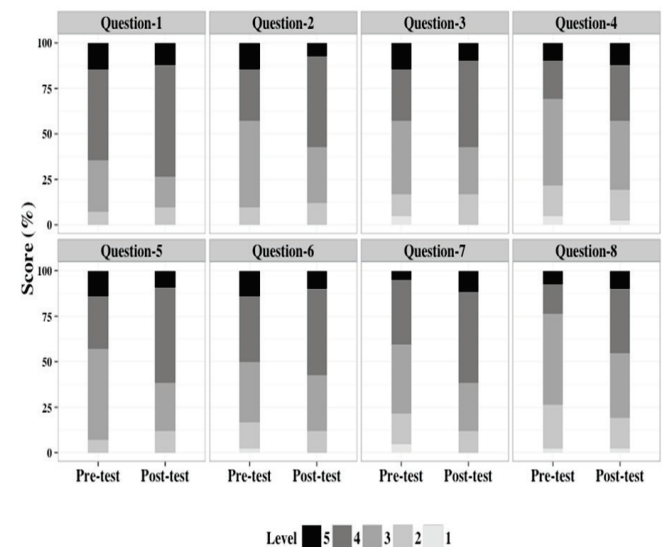


Figure 1 The Likert-scale questionnaire's results

The highlighted results of each question are as follows:

Question 1 Students are interested in learning software development. Average scores of 4 for TDD (very interested) were higher than the waterfall model by 11.9%.

Question 2 Students are interested in software development practice with exercises (The software development exercise serves as programming practice). The study found that after TDD, average scores of 4 (very interested) were higher than the waterfall model by 21.30%.

Question 3 Students are interested in software development practice with self-study (Software development exercises serve as the programming practice). From the questionnaire after TDD study, average scores of 4 (very

interested) were higher than the waterfall model by 19.05%, and average scores of 1 (not at all interested) decreased by -4.76%.

Question 4 Students are interested in learning more about software engineering by themselves (additional software engineering studies). Average scores of 5 (extremely interested) increased by 2.38%, and average scores of 4 (very interested) increased by 9.52%. Average scores of 1 (not at all interested) decreased by -2.38%.

Question 5 Students are interested in modifying the interface display from the exercise. Average scores of 4 (very interested) increased by 11.91% after using TDD. Average scores of level 1 (not at all interested) decreased by -2.38%.

Question 6 Students are interested in participating in the question & answer session in class. The results after TDD, average scores of 5 (extremely interested) were higher than the waterfall model by 2.38%, and scores of 4 (very interested) were higher than the waterfall model by 19.04%.

Question 7 Students are interested in self-learning. The results after TDD study shows that average scores of 5 (extremely interested) increased by 7.14% and scores of 4 (very interested) increased by 14.29%. Average scores of 1 (not at all interested) decreased by -4.76%.

Question 8 Students are interested in spending time working on the exercises. From the results of the questionnaire after TDD study, average scores of 4 were higher than the waterfall model by 23.81%.

In addition to the analysis with the Likert-scale questionnaire, we performed a statistical analysis with the paired Student's T-test to prove our research hypotheses.

The paired Student T-tests were analysed to compare the results of the pre-test and post-test total score of each participant (1 point denotes not at all interested and 5 denotes "extremely interested"). The results obtained from the paired Student's T-test showed that the mean scores of the post-test are statistically significantly higher than the pre-test ($p\text{-value} = 0.00365$ at a confidence level of 95%). Therefore, H_{null} was rejected, whereas $H_{\text{alternative}}$ was accepted, which means that the interest of the students who learned about TDD is different from their interest associated with the waterfall model. Based on the prior results, this may imply that the use of TDD can encourage a sample group to become more interested in software development.

4.2 Open-Ended Questions and Interviews

The answers for the open-ended questionnaires and interviews were the descriptive data, which were analysed using a coding analysis method with ATLAS.ti software (<http://atlasti.com/>). We summarized the findings as follows:

Finding 1 Effect on students' interest in using TDD in software development. From data collected from 52 questionnaires and interviews, TDD affects student interest in software development by 86%. Based on the responses of students who found that TDD affected their interest in software development, we classified the benefits of using TDD into 3 areas:

- 1) Errors found in the program. The respondents reported that TDD helped them easily check for program errors;

thus, the number of errors was reduced. Additionally, they could find program errors from the beginning of the development process.

- 2) Program quality. In terms of the quality of the code, the respondents indicated that TDD helped them reduce the complexity because refactoring methods were used during the development. Additionally, refactoring reduced the number of lines of code. Some respondents mentioned that the code they wrote was more understandable after refactoring.
- 3) The time spent using TDD for programming. The respondents felt that TDD saved time because there were fewer steps. However, some of respondents mentioned that they were more familiar with the waterfall model than with TDD.

Finding 2 The selection of software development methods between the waterfall model and TDD. Based on the analysis, we divided the selection of software development methods into 4 groups: 1) Using TDD in the exercise (73%); 2) Using the waterfall model in the exercise (15%); 3) Writing the test code and actual code without refactoring (11%); and 4) The methods are not different (1%).

Finding 3 Limitations of using TDD in software development from a student perspective. We have included several limitations that some students mentioned with regard to using TDD in the lab during and after the experiment as follows: the students thought that the developers who write the code should not write the test code; there are several steps in TDD, so additional study may be required, but they are more interested in programming; they cannot use the experimental tool (Eclipse); if they learn more about TDD, interest will increase with their confidence; they are familiar with the waterfall model for writing the code and fixing the errors at the run time; they think of TDD as only a new software development process and is not interesting as new technology; they have a better understanding of both traditional software development and the waterfall model method; and they need to reorganize their programming to use the TDD method.

4.3 Discussion

We have set all five indicators of interest and verified the indicators with confirmatory factor analysis through the Likert-scale questionnaires. The results of the factor analysis show that the questionnaires can be used to ask questions to measure the interest of the experimental groups.

Based on the questionnaires and interviews from the pre-experiment, students who used the waterfall model of software development were not interested in system testing and project testing in class. The students performed the tests by entering random data for testing purposes, but they did not test the execution process of the code internally. The students made modifications to the system when the errors were encountered at run time. Additionally, based on the responses to the questionnaires and interviews, the students understand

that using Exception try-catch in Java is one way to test a system.

From the questionnaires and the interviews about additional study, 70% of the students were interested in additional programming studies. However, the students were not interested in studying system testing and the programming process or development methods for software.

The study found that the use of TDD in class contributed to students' interest in software development. The questionnaires and interviews in pre- and post-experiment showed an increasing number of students who were interested in software development. The results of this study indicate that TDD could help encourage students' interest in software development since today's students are less interested in being software developers. The use of TDD could help encourage student interest and enables students to produce more efficient software and better quality code. This evidence paves a way to changing the learning style to produce better software developers into the industry.

These results are similar to those in previous works. For example, previous studies [16], [18], [20], [32] suggested that TDD should be used in a class to provide students with software development skills because it can help improve the quality of software development. In addition, the use of TDD shows that students make fewer programming errors [19].

5 THREATS TO VALIDITY

The threats to this study's validity are divided into *Construct*, *Internal* and *External* threats.

5.1 Construct Validity

Construct validity is a consideration if the concepts being studied are correct and all selected studies can answer the research questions. In this study, we measured student interest using questionnaires, and we studied various theories related to the definition of "interest". Nevertheless, there are various definitions of interest in the existing literature, and there is no accepted definition of interest in software engineering research. Therefore, we combined existing definitions as the indicators to formulate the questions in the questionnaire. To reduce this threat, the research questionnaire was reviewed by a software engineering expert who works in software engineering education. Additionally, the questionnaire was determined with confirmatory factor analysis, and the results suggested that the questions involved the same factor.

5.2 Internal Validity

Internal validity focuses on the results of experiments. This study is an experimental study that uses designed questionnaires and interviews. Prior to the experiment, a pilot study was conducted to check the questionnaires and interview questions in terms of whether the students could understand the exact questions. In addition, before the actual experiment, we informed the participants that the results of the questionnaire and the interview had no effect on their

scores in the studied subjects. Since the interview was semi-structured, the participants' questions might have been different. We attempted to reduce this risk by asking the interviewers to discuss and practice together. Additionally, the lessons learned from the pilot study mitigated this threat.

5.3 External Validity

External validity focuses on the generalizability of the results of this study. The experiments in the study used a sample of students in software engineering who have completed basic programming classes in OOP and Java programming. Additionally, the program we used in this study is simple. The replicated experiment using larger programs and other student fields can help mitigate these threats.

6 CONCLUSION

This study aimed to enable students to learn and use TDD to encourage their interest using experimental methods and data collection from questionnaires and interviews. The results of the experiment have proven the research hypothesis: the students who were taught to use TDD were interested in software development at a different level than those who used the waterfall model. The students who used TDD in their studies were more interested in software development. Based on this study, we believe that the results of the research will help investigators in software engineering encourage interest in software development. Consequently, it also focused on software development methods and the emotional attitudes of software developers. The study is beneficial to educational personnel, especially in software engineering education, to change teaching methods and motivate students to become more interested in software development. Finally, it provides empirical evidence in software engineering regarding the use of TDD in the classroom, including constraints and recommendations for other researchers or educational personnel using TDD in teaching and learning.

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

5 REFERENCES

- [1] Kaczmarczyk, L. & Dopplick, R. (2014). ACM Report: Preparing Students for Computing Workforce Needs in the U.S. *ACM SIGCSE Bulletin*, 46(2), 8-8.
<https://doi.org/10.1145/2620668.2620675>
- [2] <https://www.nap.edu/catalog/24926/assessing-and-responding-to-the-growth-of-computer-science-undergraduate-enrollments>
- [3] <https://insights.stackoverflow.com/survey/2019>
- [4] <http://www.nationmultimedia.com/detail/Economy/30344732>
- [5] Royce, W. W. (1987). Managing the development of large software systems: concepts and techniques. In *Proceedings of*

- the 9th international conference on Software Engineering (pp. 328-338)
- [6] Edwards, S. H. (2004, March). Using software testing to move students from trial-and-error to reflection-in-action. *ACM SIGCSE Bulletin*, 36(1), 26-30. <https://doi.org/10.1145/1028174.971312>
 - [7] Piaget, J. (1977). *The role of action in the development of thinking*. In Knowledge and development. Springer, Boston, MA, 17-42.
 - [8] Cakir, M. (2008). Constructivist approaches to learning in science and their implications for science pedagogy: A literature review. *International journal of environmental and science education*, 3(4), 193-206. https://doi.org/10.1007/978-1-4684-2547-5_2
 - [9] Beck, K. (2003). *Test-driven development: by example*. Addison-Wesley Professional.
 - [10] Karac, I. & Turhan, B. (2018). What Do We (Really) Know about Test-Driven Development? *IEEE Software*, 35(4), 81-85. <https://doi.org/10.1109/MS.2018.2801554>
 - [11] Desai, C., Janzen, D., & Savage, K. (2008). A survey of evidence for test-driven development in academia. *ACM SIGCSE Bulletin*, 40(2), 97-101. <https://doi.org/10.1145/1383602.1383644>
 - [12] Munir, H., Moayyed, M., & Petersen, K. (2014). Considering rigor and relevance when evaluating test driven development: A systematic review. *Information and Software Technology*, 56(4), 375-394. <https://doi.org/10.1016/j.infsof.2014.01.002>
 - [13] Scanniello, G., Romano, S., Fucci, D., Turhan, B., & Juristo, N. (2016, April). Students' and professionals' perceptions of test-driven development: a focus group study. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, 1422-1427. <https://doi.org/10.1145/2851613.2851778>
 - [14] Mumtaz, H., Alshayeb, M., Mahmood, S., & Niazi, M. (2018). An empirical study to improve software security through the application of code refactoring. *Information and Software Technology*, 96, 112-125. <https://doi.org/10.1016/j.infsof.2017.11.010>
 - [15] Tosun, A., Dieste, O., Fucci, D., Vegas, S., Turhan, B., Erdogmus, H., & Juristo, N. (2017). An industry experiment on the effects of test-driven development on external quality and productivity. *Empirical Software Engineering*, 22(6), 2763-2805. <https://doi.org/10.1007/s10664-016-9490-0>
 - [16] Janzen, D. & Saiedian, H. (2005). Test-driven development concepts, taxonomy, and future direction. *Computer*, 38(9), 43-50. <https://doi.org/10.1109/MC.2005.314>
 - [17] Janzen, D., & Saiedian, H. (2008). Test-driven learning in early programming courses. *ACM SIGCSE Bulletin*, 40(1), 532-536. <https://doi.org/10.1145/1352322.1352315>
 - [18] Pancur, M., Ciglaric, M., Trampus, M., & Vidmar, T. (2003, Septe). Towards empirical evaluation of test-driven development in a university environment. In *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, 2, 83-86.
 - [19] Edwards, S. H. (2003, August). Using test-driven development in the classroom: Providing students with automatic, concrete feedback on performance. In *Proceedings of the international conference on education and information systems: technologies and applications EISTA* (Vol. 3).
 - [20] Buffardi, K. (2012, September). Understanding and persuading adherence to test-driven development. In *Proceedings of the ninth annual international conference on International computing education research*, 155-156. <https://doi.org/10.1145/2361276.2361308>
 - [21] Kropp, M., Meier, A., Anslow, C., & Biddle, R. (2018, June). Satisfaction, practices, and influences in agile software development. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering*, 112-121. <https://doi.org/10.1145/3210459.3210470>
 - [22] Bissi, W., Neto, A. G. S. S., & Emer, M. C. F. P. (2016). The effects of test driven development on internal quality, external quality and productivity: A systematic review. *Information and Software Technology*, 74, 45-54. <https://doi.org/10.1016/j.infsof.2016.02.004>
 - [23] Munir, H., Wnuk, K., Petersen, K., & Moayyed, M. (2014, May). An experimental evaluation of test driven development vs. test-last development with industry professionals. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (p. 50). <https://doi.org/10.1145/2601248.2601267>
 - [24] Müller, M. M. & Höfer, A. (2007). The effect of experience on the test-driven development process. *Empirical Software Engineering*, 12(6), 593-615. <https://doi.org/10.1007/s10664-007-9048-2>
 - [25] Mäkinen, S. & Münch, J. (2014, January). Effects of test-driven development: A comparative analysis of empirical studies. In *International Conference on Software Quality*, Springer, Cham, 155-169. https://doi.org/10.1007/978-3-319-03602-1_10
 - [26] Causevic, A., Sundmark, D., & Punnekkat, S. (2011, March). Factors limiting industrial adoption of test driven development: A systematic review. In *2011 Fourth IEEE International Conference on Software Testing, Verification and Validation*, 337-346. <https://doi.org/10.1109/ICST.2011.19>
 - [27] Dwyer, K. K. (1993). Using the 4MAT System Learning Styles Model to Teach Persuasive Speaking in the Basic Speech Course.
 - [28] Charters, W. W. & Good, C. V. (1945). The Dictionary of Education. *The Phi Delta Kappan*, 27(1), 5-7.
 - [29] Fenton, S. (2018). Automated Testing. In *Pro TypeScript*. Apress, Berkeley, CA, 245-256. https://doi.org/10.1007/978-1-4842-3249-1_10
 - [30] Fox, R. J. (2010). *Confirmatory factor analysis*. Wiley International Encyclopedia of Marketing. <https://doi.org/10.1002/9781444316568.wiem02060>
 - [31] Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141-151. <https://doi.org/10.1177/001316446002000116>
 - [32] Romano, S., Fucci, D., Baldassarre, M. T., Caivano, D., & Scanniello, G. (2019). An Empirical Assessment on Affective Reactions of Novice Developers when Applying Test-Driven Development. *arXiv preprint arXiv:1907.12290*. https://doi.org/10.1007/978-3-030-35333-9_1

Authors' contacts:

Aziz NANTHAAMORNPHONG

College of Computing, Prince of Songkla University,
80 Moo1 Vichitsongkram Road, Kathu, Phuket, 83120, Thailand
aziz.n@phuket.psu.ac.th

Stephane BRESSAN

School of Computing, National University of Singapore,
21 Lower Kent Ridge Road, Singapore 119077
steph@comp.nus.edu.sg

A GRAPH-BASED ALGORITHM FOR INTERPERSONAL TIES CLUSTERING IN SIGNED NETWORKS

Sumalee SANGAMUANG

Abstract: Social ties are formed as a result of interactions and individual preferences of the people in a social network. There are two opposite types which are interpreted as friendship vs. enmity or trust vs. distrust between people. The aforementioned social network structure can be represented by a signed graph, where people are the graph's vertices and their interactions are graph's edges. The edges can be positive and negative signs. To determine trustworthiness, this paper considers the problem of a signed graph partitioning with minimizing the sum of the negative edge's weight and balanced size of its clusters. An efficient algorithm to solve such a problem is proposed. The experimental results show that the proposed algorithm outperforms in terms of the execution times and the accuracy within the given bounds.

Keywords: graph partition; min-cuts; signed graphs; social networks; social ties

1 INTRODUCTION

Recently, the social internet of things (SIoT) has been proposed and is the subject of a rapidly increasing research effort [1], [2] and [3]. The structure of SIoT has been formulated in [4] based on the notion of social relationships among the IoT objects. The aforementioned relationships can be represented by a signed graph where the vertices are the IoT objects and their interactions are the edges. The problem of service selection was discussed in [5] based on social influence. Given the set of IoT nodes providing the needed service, every node, after interaction with another node, gives a rating to the node based on the service it provides. All nodes interact among themselves based on a social network structure. Each object has to be provided with the best and most trusted reliable node.

In a social network, social ties are formed as a result of interactions and individual preferences of the people in a social network. One person may change to another in the causes of behavior related to reliability. There is a level of trustworthiness in a social network that can be determined by the PageRank [6-8]. Thus, the level of trustworthiness is considered from the values of PageRank incoming links. For real-world scenarios, there are two opposite types, i.e. there is trust versus distrust between the relationships. Thus, the edges should be the positive and negative signs.

Fig. 1 shows an example of a signed graph that represents the social relationships among IoT objects in a real-world scenario. In this figure, trust is represented by a straight line, while distrust is represented by a dotted line. The level of trustworthiness is represented by the edge's weight. If the edge's weight is small, there is not enough trustworthiness.

On the other hand, if the edge's weight is large, there are high levels of trustworthiness. The number of IoT devices and services is steadily increasing, and given the connectivity and reliability of a large signed chart of centralization systems, it is not efficient. A distributed system is a viable choice. In [5], a large-scale service selection problem was

considered. Based on such a problem, the large signed graph has to be partitioned into smaller clusters for reducing execution times.

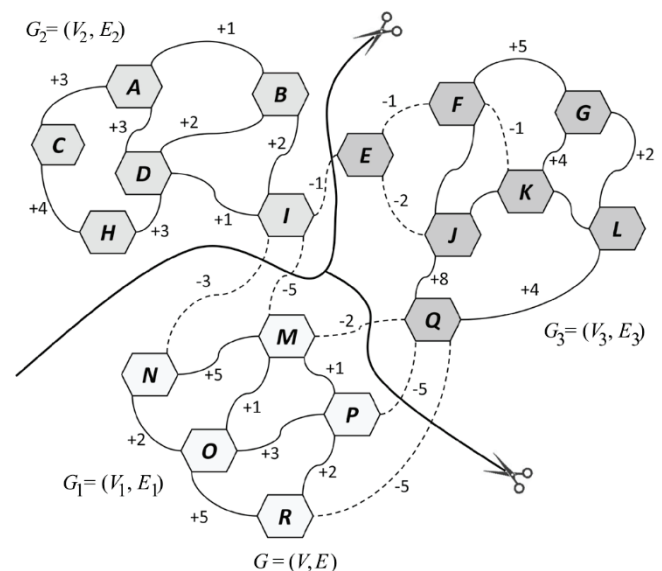


Figure 1 An example of the minimal partition of the signed graph

The set of clusters is called the partition of the signed graph. For each input graph, there is a number of possible partitions. This paper considers an optimization problem to find approximately the same cluster size that the sum of the weight of the negative external-edge's is minimized while the sum of the positive external-edges is less than the given constant. A graph-based algorithm for interpersonal ties clustering in signed networks is proposed to solve the aforementioned problem. The example of the optimal partition is showed in Fig. 1. Let integers $k = 3$ and $\omega = 5$ and $G = (V, E)$ be a signed graph. The signed graph is partitioned into its minimal partition $P_c = \{G_1, G_2, G_3\}$ where the sum of negative external edge's weight is -21 , as a minimized edge's weight, and the sum of positive external edge's weight is 0 (less than a given constant ω).

2 PROBLEM DEFINITION INTRODUCTION

In this section, the notations and problem statement are defined as follows.

Definition 1 (Signed Graph): Let $G = (V, E)$ be a signed graph which consists of n vertices. The vertices' set $V = \{v_1, v_2, v_3, v_4, \dots, v_n\}$, and the edges' set $E = \{(v_i, v_j, w_{ij}) \mid i \neq j, w_{ij} \in I\}$.

Definition 2 (Cluster): Given $G = (V, E)$, a graph $G_c = (V_c, E_c)$ is a cluster of G where $V_c \subseteq V$ and $E_c \subseteq E$.

Definition 3 (Partition): Given $G = (V, E)$, a set of its clusters $P_c = \{G_1, G_2, G_3, G_4, \dots, G_c\}$ is called a partition of G that c is the number of clusters in the partition.

Definition 4 (Internal Edge): Given $G = (V, E)$, and let P_c be its partition. For each cluster G_i in P_c , a pair $(v_i, v_j, w_{ij}) \in E_c$ is called an internal-edge. There are two types of an internal-edge, i.e., a negative internal-edge and a positive internal-edge. A set of a positive internal-edge is $E^+ = \{(v^+, v^+, w^+) \mid w^+ \in I^+\}$ while a set of a negative internal-edge is $E^- = \{(v^-, v^-, w^-) \mid w^- \in I^-\}$ that $E^+ \cup E^-$.

Definition 5 (External Edge): Given $G = (V, E)$, and let P_c be its partition. For each cluster G_i in P_c , a pair $(v_i, v_j, w_{ij}) \in E \setminus \bigcup_{c=1}^k E_c$ is called an external-edge.

Definition 6 (Cut): Given $G = (V, E)$, and let P_c be its partition. For each cluster G_i in P_c , a set of the external-edges is called a cut C which consists of a set of positive external-edges $C^+ = \{(v_i^+, v_j^+, w_{ij}) \mid w_{ij} \in I^+\}$ and a set of negative external edges $C^- = \{(v_i^-, v_j^-, w_{ij}) \mid w_{ij} \in I^-\}$.

This paper considers an optimization problem to find k a roughly equal size of the clusters that the sum of negative external-edge's weight is minimized, while the sum of positive external-edges is less than the given constant. For the problem instance, it is given $G = (V, E)$ and positive integers k and ω . The problem returns P_c which consists of k roughly equal size of the clusters, meanwhile, the sum of the edge's weight w_{ij} in C^- is minimized, and the sum of the edge's weight w_{ij} in C^+ is less than ω . This problem is the under category NP-hard by reducing to the min-cut max-flow problem.

3 THE ALGORITHM

In [9] and [10], an algorithm for partitioning a graph into balanced clusters with a minimal size of cuts was proposed. Based on the aforementioned algorithm, this section presents a heuristic algorithm to find balanced clusters whose sum of the negative external-edge's weight is minimized, while the sum of positive external-edges is less than a given constant.

Algorithm 1 separates a signed graph $G = (V, E)$ into its partition P_c with a satisfied condition at line 1. The sum of all edge's weights in C^+ is less than ω . The input graph is recurrently separating itself into a smaller cluster G_c until the size of P_c is k (line 9 to line 13).

In Line 9, the set of vertices of the input graph are separated into two subsets, $X \in V$ and $Y \in V$, with an equal size by using the Algorithm 2. Line 10 and line 11 create clusters from x_i and y_i . Thus, both clusters are equal to the

number of vertices and are contained within a partition of the input graph. This partition is called an exact bisection of the input graph. Since both clusters half-separate themselves at line 12 and line 13, the Algorithm 1 runs in $O(|V| \log |V|)$ time.

Algorithm 1 Find the Cluster

Require: A signed graph $G = (V, E)$ and constants k, ω .

Ensure: A partition P_c .

```

1: if  $f(C^+) < \omega$  &  $|P_c| < k$  then
2:   if  $P_c = \emptyset$  then
3:      $P_c \neq \emptyset$ 
4:   else
5:      $P_c \neq G_c$ 
6:   end if
7:   return  $P_c$ 
8: else
9:   find the cut-set  $C_{V_i, V_j}$  using Algorithm 2
10:  create subgraphs  $G_i$  from  $V_i$ 
11:  create subgraphs  $G_j$  from  $V_j$ 
12:   $G_c \leftarrow \text{FindCluster}(G_i, \omega)$ 
13:   $G_c \leftarrow \text{FindCluster}(G_j, \omega)$ 
14: end if

```

Algorithm 2 is formulated based on the well-known heuristic algorithm which is called Kernighan-Lin [11]. The set of vertices V is separated into two disjoint subsets X and Y of equal or nearly equal size which minimizes the cost of crossing edges (the sum of the edge's weight). However, the Kernighan-Lin algorithm is limited with the negative edges of the signed graph.

For determining the cost, let C_I be the internal cost of $x \in X$, which is the sum of the edge's weight w_x between x and other nodes in X , i.e., $\sum_{w_x \in X} (w_x)^2$. Let C_E be the external cost of $x \in X$ which is the sum of the edge's weight $w_{x,y}$ between the nodes $x \in X$ and $y \in Y$, i.e., $\sum_{w_{x,y} \in X} (w_{x,y})^2$.

Therefore, the difference of the internal cost and the external cost is formulated as follows:

$$D_x = C_E - C_I \quad (1)$$

Moreover, the cost for an interchange between x and y is

$$g_{x,y} = D_x + C_y - 2C_{x,y} \quad (2)$$

The term $C_{x,y}$ is the cost of the possible edge between x and y .

The algorithm tries to find an optimal series of interchange operations between the elements, minimizing the sum of the external edge's weight between x and y , which maximizes $g_{x,y}$ and then an optimal bisection of the input graph will be produced when the algorithm is finished. Algorithm 2 runs in $O(|V| \log |V|)$ time.

Algorithm 2 Find the max cut**Require:** Given $G = (V, E)$.**Ensure:** The max-cut of G .

```

1: Randomly half-separates the set  $V$  into sets  $X$  and  $Y$ 
2: while  $g_{\max} \leq 0$  do
3:   Compute  $D$  values for all  $x_i \in X$  and  $y_i \in Y$ 
4:   Let  $g$ ,  $X^*_i$  and  $Y^*_i$  be empty lists
5:   Let  $k \leftarrow |V|/2$ 
6:   for  $i \leftarrow 1$  to  $k$  do
7:     Find cost for the interchange between  $x_i$  and  $y_i$  by
       Eq. (2)
8:     Remove  $x_i$  and  $y_i$  from further consideration in this
       step
9:     Add  $g$  to  $g[i]$ ,  $x_i$  to  $X^*[i]$ , and  $y_i$  to  $Y^*[i]$ 
10:    Update  $D$  values for the elements of  $X^*[i] = X \setminus x_i$ 
       and  $Y^*[i] = Y \setminus y_i$ 
11:    Find index  $i$  of  $(x_i, y_i)$  which  $g_{\max} = \max_{i=1}^k g[i]$ 
12:  end for
13: end while
14: if  $g_{\max} > 0$  then
15:   Swap  $X^*[i]$  with  $Y^*[i]$ 
16:   Set  $X = X^*$  and  $Y = Y^*$ 
17: end if

```

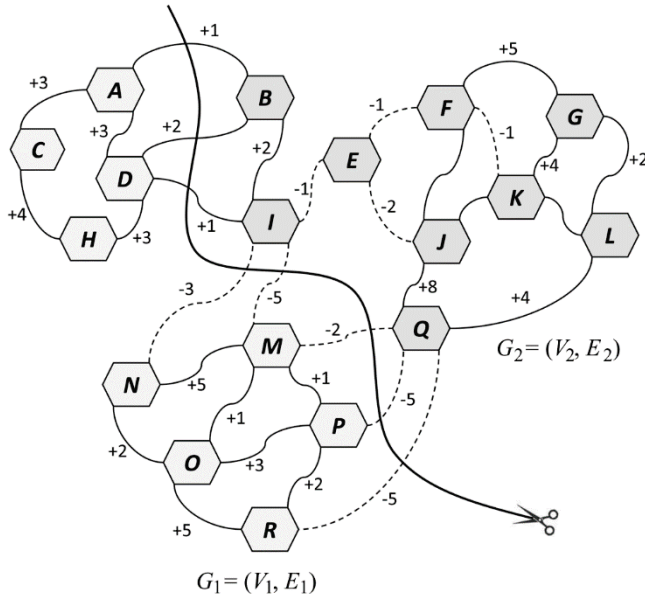
**Figure 2** An example of the clusters of the input graph by using the Algorithm 2

Fig. 2 shows an example of the clusters of the input graph by using the Algorithm 2. Let $E = 5$, the input graph is separated into its partition which consists of the cluster G_i and the cluster G_j of equal sizes. Based on the heuristic approach, there are three positive external edges whose sum of the edge's weight is four. This partition can be acceptable because the sum of all edge's weight in C^+ is less than five.

4 EXPERIMENT EVALUATION

This section evaluates the performances, i.e., the partition's execution times and the accuracy. The experimental results show the performances of the proposed algorithm and MINOS solver by measuring the execution times and the accuracy to the partition. Moreover, the performances of the proposed algorithm on a real-world large-scale signed social network are evaluated.

The experiments use the datasets from Slashdot [12], a signed social network that contains friend/foe links between the Slashdot's users. This dataset was obtained in November 2008, November 2009 and February 2009 and it contains 77,350 users with 516,575 relationships, 81,871 users with 545,671 relationships and 82,144 users with 549,202 relationships.

Table 1 Performances in terms of execution times

$ V $	Execution Times (min.)	
	Minos Solver	Proposed Algorithm
40	1.07	0.21
80	5.24	0.39
120	13.11	0.41
160	89.13	0.56
200	245.43	0.69

Table 2 Performances in terms of accuracy

$ V $	Cuts	
	Minos Solver	Proposed Algorithm
40	16	100
80	21	152
120	23	203
160	38	240
200	52	298

Tab. 1 shows the execution times of the proposed algorithm and MINOS solver. If the graph sizes are increased, the execution times of MINOS solver will rapidly increase. On the other hand, the execution times of the proposed algorithm are less than one minute. For example, at the graph size of 200 vertices, the execution time of the MINOS solver is 245.43 minutes, while the execution time of the proposed algorithm is only 0.69 minute.

Tab. 2 shows the sizes of cuts from the partitions using the proposed algorithm and MINOS solver. The sizes of cuts from the partitions using the proposed algorithm are always higher. However, these solutions are an acceptable region in the application for a social network analysis.

Tab. 1 and Tab. 2 present trade-off between the execution times and the accuracy. Since the MINOS optimizer is limited with solving very large-scale problems, the proposed algorithm is a viable choice. However, the execution times of the proposed algorithm had to be traded by its accuracy.

Tab. 3 shows the performances of the proposed algorithm on a real-world large-scale signed social networks, i.e. Slashdot in November 2008, Slashdot in November 2009 and Slashdot in February 2009. The experimental results show that the cuts of partitions from Slashdot in November 2008 are 1.14 percent, Slashdot in November 2009 are 1.93

percent and Slashdot in February 2009 are 2.17 percent, meanwhile, these execution times are less than one hour.

Table 3 Performances in terms of the proposed algorithm on large-scale networks

Signed Networks	Execution Times (min.)	Cuts
Slashdot (November 2008)	19.17	5873
Slashdot (November 2009)	37.62	10520
Slashdot (February 2009)	51.18	11905

For measuring the ranking accuracy, the Spearman's rank correlation coefficient [13] has been used for determining the degrees of the ranking's error. It is defined as $F(\sigma_1, \sigma_2) = \sum |\sigma_1(i) - \sigma_2(i)|$, where $\sigma_1(i)$ and $\sigma_2(i)$ are the rank of the vertex i in the centralized vertices ranking and distributed vertices ranking, respectively. This value indicates the error from the distributed vertices ranking, higher Spearman's footrule distance indicates higher degrees of the ranking's error.

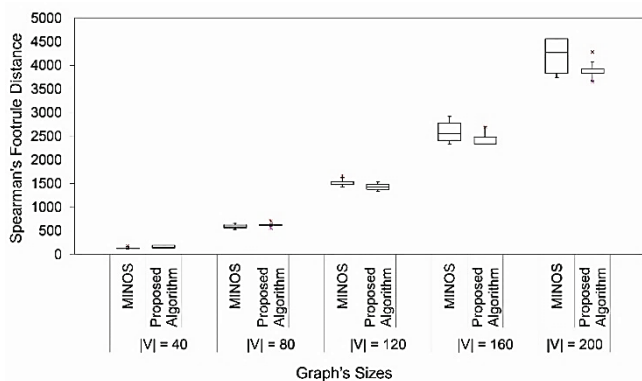


Figure 3 The ranking accuracy of the proposed algorithm compared with the MINOS solver

Fig. 3 shows the ranking accuracy of the proposed algorithm compared with the MINOS solver. The X-axis is the graph size in the ranges from 40 to 200, while the Y-axis is the Spearman's footrule distance. The proposed algorithm can perform faster when compared with the MINOS solver, while the accuracy is still acceptable for the vertices ranking on social networks.

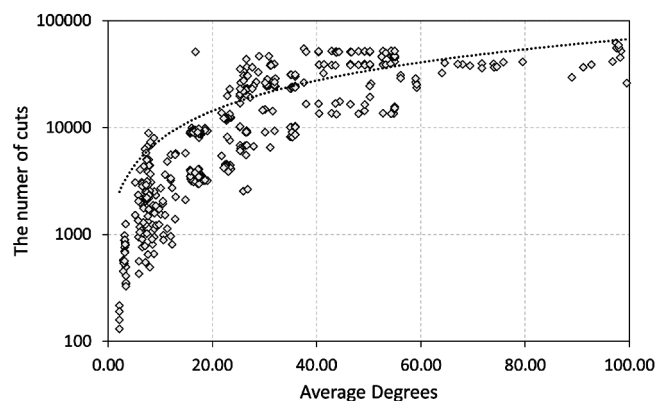


Figure 4 The cuts of the proposed algorithm

Fig. 4 shows impacts of the average degrees on the size in the ranges from 0 to 100, while the Y-axis is the size of the

cuts. This experiment uses 150 different randomly selected subgraphs of Slashdot in November 2008 to create 150 balanced partitions.

In this figure, if the average degrees are increased, the sizes of cuts will rapidly increase until the average degree is 30. However, the graph sizes gradually increase when the average degree is over 30. Likewise, Fig. 5 shows the impacts of the average degrees on execution times. It can clearly be seen that the execution times of the proposed algorithm are the logarithm of the average degrees of the graph.

Based on the experimental results, the proposed algorithm is very efficient for partitioning a large-scale signed social network.

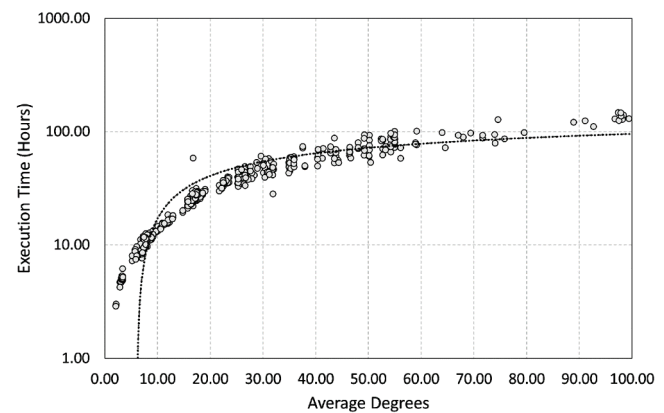


Figure 5 The execution times of the proposed algorithm

5 CONCLUSIONS

This paper considers an optimization problem to find roughly equal size of the clusters that the sum of negative external-edge's weight is minimized, while the sum of the positive external-edges is less than a given constant. A graph-based algorithm for interpersonal ties clustering in signed networks is proposed to solve the aforementioned problem. The experimental results show that the proposed algorithm can perform faster when compared to the MINOS solver, while the accuracy is still acceptable for the vertices ranking on social networks. Moreover, experimental results show that the number of cuts has an upper bound which will be considered as an approximation bound in future works.

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

6 REFERENCES

- [1] Iera, A., Morabito, G., & Atzori, L. (2015, March). The social internet of things. In *2015 IEEE International Conference on Cloud Engineering*, 1-1. <https://doi.org/10.1109/IC2E.2015.68>
- [2] Nitti, M., Atzori, L., & Cvijikj, I. P. (2015, June). Friendship selection in the social internet of things: Challenges and

- possible strategies. *IEEE Internet of Things Journal*, 2(3), 240-247. <https://doi.org/10.1109/JIOT.2014.2384734>
- [3] Ortiz, A. M., Hussein, D., Park, S., Han, S. N., & Crespi, N. (2014, June). The cluster between internet of things and social networks: Review and research challenges. *IEEE Internet of Things Journal*, 1(3), 206-215. <https://doi.org/10.1109/JIOT.2014.2318835>
- [4] Atzori, L., Iera, A., & Morabito, G. (2011, November). Siot: Giving a social structure to the internet of things. *IEEE Communications Letters*, 15(11), 1193-1195. <https://doi.org/10.1109/LCOMM.2011.090911.111340>
- [5] Nizamkari, N. S. (2017, Jan). A graph-based trust-enhanced recommender system for service selection in IoT. In *2017 International Conference on Inventive Systems and Control (ICISC)*, 1-5. <https://doi.org/10.1109/ICISC.2017.8068714>
- [6] Chen, S. & He, K. (2015, Dec). Influence maximization on signed social networks with integrated PageRank. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 289-292. <https://doi.org/10.1109/SmartCity.2015.86>
- [7] Hajian, B. & White, T. (2011, Oct). Modelling influence in a social network: Metrics and evaluation. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 497-500. <https://doi.org/10.1109/PASSAT/SocialCom.2011.118>
- [8] Wu, J. (2017, March). Realtime personalized pagerank query for social network search. In *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 647-651. <https://doi.org/10.1109/IAEAC.2017.8054096>
- [9] Feige, U. & Krauthgamer, R. (2002, Apr). A polylogarithmic approximation of the minimum bisection. *SIAM J. Comput.*, 31(4), 1090-1118. <https://doi.org/10.1137/S0097539701387660>
- [10] Andreev, K. & Racke, H. (2006, November). Balanced graph partitioning. *Theory of Computing Systems*, 39, 929-939. <https://doi.org/10.1007/s00224-006-1350-7>
- [11] Kernighan, B. W. & Lin, S. (1970, Feb). An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 49(2), 291-307. <https://doi.org/10.1002/j.1538-7305.1970.tb01770.x>
- [12] Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '10*. New York, NY, USA: ACM, 1361-1370. <https://doi.org/10.1145/1753326.1753532>
- [13] Fagin, R., Kumar, R., & Sivakumar, D. (2003). Comparing top k lists. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 28-36. <https://doi.org/10.1137/S0895480102412856>

Author's contacts:

Sumalee SANGAMUANG, PhD
Chiang Mai University,
College of Arts, Media and Technology,
239 Huaykaew Rd., Suthep, Muang,
Chiang Mai, 50200 Thailand
sumalee.sa@cmu.ac.th

EVALUATING BIOMETRICS FINGERPRINT TEMPLATE PROTECTION FOR AN EMERGENCY SITUATION

Ei Ei MON, Sangsuree VASUPONGAYYA, Montri KARNJANADECHA, Touchai ANGCHUAN

Abstract: Biometric template protection approaches have been developed to secure the biometric templates against image reconstruction on the stored templates. Two cancellable fingerprint template protection approaches namely minutiae-based bit-string cancellable fingerprint template and modified minutiae-based bit-string cancellable fingerprint template, are selected to be evaluated. Both approaches include the geometric information of the fingerprint into the extracted minutiae. Six modified fingerprint data sets are derived from the original fingerprint images in FVC2002DB1_B and FVC2002DB2_B by conducting the rotation and changing the quality of original fingerprint images according to the environment conditions during an emergency situation such as wet or dry fingers and disoriented angle of fingerprint images. The experimental results show that the modified minutiae-based bit-string cancellable fingerprint template performs well on all conditions during an emergency situation by achieving the matching accuracy between 83% and 100% on FVC2002DB1_B data set and between 99% and 100% on FVC2002DB2_B data set.

Keywords: biometric-based user authentication; biometric cryptosystem; cancellable biometrics

1 INTRODUCTION

Biometric-based authentication system uses unique and measurable biological characteristics of an individual to verify a person [1]. Biometric-based authentication is more convenient than other user authentication techniques based on passwords or smartcards. Passwords can be forgotten and smartcards must be carried. Biometric-based authentication systems are used in many applications such as law enforcements, healthcare, banking, access controls and smart phones. Fingerprint is a cheap and convenient way to identify someone because fingerprint scanners are small and cheap in comparison with other biometric capturing devices [1, 2]. However, a biometric-based authentication system has a serious drawback on privacy. Under a traditional fingerprint-based authentication system, an unknown original fingerprint image can be reconstructed from a stored fingerprint template [3, 4]. Therefore, biometric template protection (BTP) methods have been used for the security of fingerprint templates. Fig. 1 shows the fingerprint-based authentication system with the template protection.

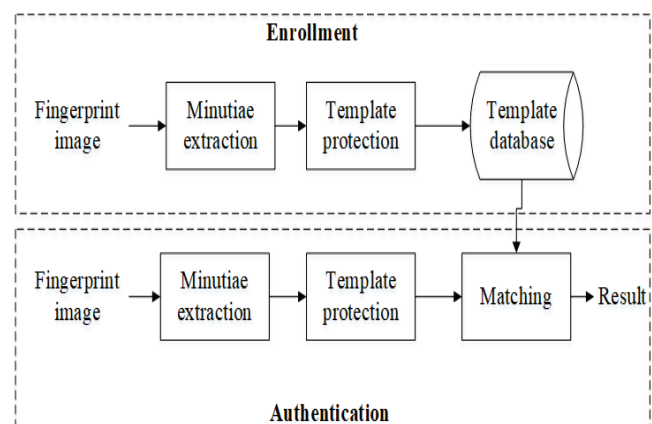


Figure 1 Fingerprint-based authentication system with a template protection

The biometric template protection schemes are commonly categorized as feature transformation (or Cancellable Biometrics (CB)) and biometric cryptosystems (BCS) as shown in Fig. 2. In this work, two CB techniques are evaluated under an emergency situation for personal health record systems. One of the challenges in fingerprint template protections is the performance degradation after performing a template protection. This issue is amplified when the user is unconscious because the quality of the fingerprint image might be affected by many environment conditions and human errors. During an emergency, the emergency service person is the main person to collect the fingerprint images from the victim. Thus, the fingerprint image quality might be affected by environment such as wet or dry fingers and the orientation of the finger during the fingerprint collection process.

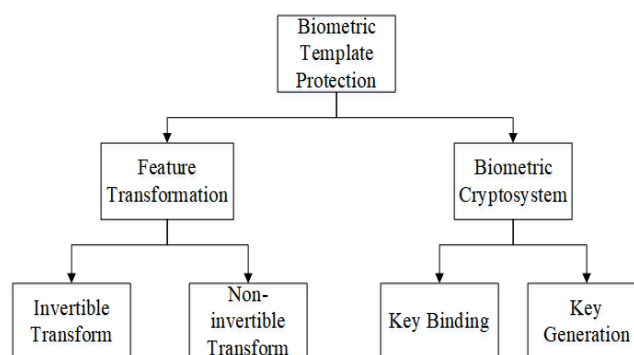


Figure 2 Biometric template protection techniques [3]

In this work, two CB techniques are evaluated with various fingerprint quality generated according to the environment conditions during an emergency situation such as wet or dry fingers and disoriented angle of fingerprint images. The result of this work will be further used for proposing an improved fingerprint feature transformation

approach for personal health record users during an emergency situation.

The remaining sections of this paper are organized as follows. Section 2 provides information on biometric template protection. The experimental settings are described in Section 3. Results and discussions are given in Section 4 while the conclusion is shown in Section 5.

2 BIOMETRIC TEMPLATE PROTECTION

The fingerprint-based authentication system consists of two main processes, including enrolment and authentication [2]. For the enrolment process, a fingerprint image is captured and features are extracted from the fingerprint image. The biometric template protection algorithm is performed on the fingerprint templates and stored in the system database as the protected templates. Depending on whether the biometric system is applied for the identification or the verification process, the authentication is performed differently. The verification system confirms the person's identity by performing a one-to-one comparison between the queried fingerprint and the stored fingerprint template. The identification system performs a one-to-many comparison by searching the entire system database.

According to the international standard ISO/IEC 24745: biometric information protection, the biometric template protection schemes must meet these properties: irreversibility, unlinkability, revocability and performance [5]. The irreversibility means the reconstruction from the protected biometric template to the original image or the biometric sample, should be difficult. The unlinkability refers to the link between the protected biometric template from one application to another application should not be possible. The revocability means that the new template can be revoked and re-issued like a password in an event of the compromising template information. The performance property means that the template protection method should not degrade the recognition performance across a number of applications.

The feature transformation consists of the distortion or the transformation of the biometric features based on the transformation function [3]. According to the characteristics of the transformation function, this category can be further divided into the invertible transform (salting) and the noninvertible transform. Salting is one of the template protections approaches. The user specific extra information is added to the original biometric features using an invertible function with a key or a password. The user needs to remember the key and presents it during the authentication. Therefore, the security of the salting is relied on the protection of the key. The template will be compromised when the key is stolen. The main advantage of the salting is the incorporation of the additional information into the biometric template in the form of keys. It makes difficult for an imposter to guess the template. The main drawback of the salting approach is that the security of this scheme relies upon the secrecy of the key. In the noninvertible transform, a one-way function is applied on the biometric features. Unlike the salting approach, the noninvertible function does not rely on

the secrecy of the keys. The difficulty of reversing the transformation function is the major protection of the biometric data. The protected template can be revoked and issued a new one by changing the transformation parameter in the case of compromising.

Table 1 Existing cancellable fingerprint template generation methods

No	Reference	Template Protection Approach	Pre-processing			Feature				Matching			
			Segmentation	Enhancement	Binarization	Minutiae	Singular point	Texture	Other	Correlation	Vector	Bit-String	Other
1	[5]	CB	(P-MCC)			✓			✓				✓
2	[6]	CB	(bit-string)			✓			✓		✓		
3	[8]	CB	(MCC)			✓					✓		
4	[10]	CB	(P-MCC)			✓					✓		
5	[13]	CB				✓						✓	
6	[14]	CB				✓						✓	
7	[15]	CB	✓	✓	✓	✓	✓						✓
8	[16]	CB	✓	✓	✓	✓						✓	
9	[17]	CB	✓	✓	✓	✓							✓
10	[20]	CB		✓	✓	✓					✓		
11	[11]	CB				✓						✓	
12	[18]	CB		✓	✓			✓		✓			
13	[23]	CB		✓	✓	✓					✓		
14	[7]	CB	VeriFinger			✓					✓		
15	[12]	CB	VeriFinger			✓							✓
16	[19]	CB	VeriFinger 6.0			✓	✓						✓
17	[21]	CB	VeriFinger			✓					✓		
18	[22]	CB	VeriFinger 6.0			✓					✓		

Template protection of this work is a cancellable fingerprint template generation approach. The literature review of cancellable fingerprint template generation approaches is described in Tab. 1. Researchers used different types of fingerprint features. Some approaches extracted not only minutiae but also singular point and other features [5, 6, 15, 19]. The resulted template representation was also different. Some methods performed the correlation-based matching [18]. The protected fingerprint templates were generated as a complex vector [6-8, 10, 12, 21-23] or a binary bit-string [11, 13, 14, 16]. In the design of cancellable fingerprint templates, some researchers focused on the template generation from the minutia information [9, 13, 16, 17, 20] while some researchers spent more effort on the design of non-invertible transformations [6-8, 10-11]. The protected templates in related works were generated in a binary form because of its simplicity in the feature representation. The design of non-invertible transformation functions in the related methods used a permutation process.

3 EXPERIMENTAL SETTINGS

Two cancellable fingerprint templates (CB) namely minutiae-based bit-strings CB and modified minutiae-based bit-strings CB are evaluated in this work. Fingerprint Verification Competitions (FVC) databases were used for evaluating both CB approaches. However, the fingerprints are modified to mimic the conditions of emergency situations described previously.







3.1 FVC Databases

The commonly used fingerprint evaluating databases are Fingerprint Verification Competitions (FVC) databases such as FVC2000, FVC2002, FVC2004 and FVC2006 [24-27]. Each FVC database is composed of four sub-databases DB1, DB2, DB3 and DB4. DB1, DB2 and DB3 were collected using optical, capacitive and thermal sensors, respectively. DB4 was created using Synthetic Fingerprint Generator (SFinGe). Each sub-database includes two subsets A and B. For three data sets (DB1, DB2 and DB3), subset A contains 100 fingers with 8 impressions (800 images) and subset B contains 10 fingers with 8 impressions (80 images). Fingers from set B have been available to allow a parameter tuning process and the benchmark is then tested by the fingers from set A. In FVC2000, the acquisition conditions were different for each database [24]. In FVC2002, the acquisition conditions were the same for each database [25]. Interleaved acquisition of different fingers causes differences in the finger placement. FVC2004 and FVC2006 databases were collected with the aim of creating a more difficult benchmark [26, 27].

3.2 Modified FVC Databases

During the emergency situation, the victim fingerprint cannot be placed in the same direction with the enrolled (registered) image. The collected fingerprint images can have variations such as wet finger, dry finger and the image acquisition style. Therefore, the input fingerprint images are modified in our work as proposed in [28] to create these three conditions. The fingerprint image brightness is changed to reduce the brightness by 70% (denoted B (-70)) and to increase the brightness by 35% (denoted B (+35)). The fingerprint image is rotated clockwise by 5 degrees (denoted CW5) and the fingerprint image is rotated clockwise by 10 degrees (denoted CW10). The fingerprint image is rotated counter clockwise by 5 degrees (denoted CCW5) and the fingerprint image is rotated counter clockwise by 10 degrees (denoted CCW10). Tab. 2 shows the samples modified fingerprint image used in this work.

Table 2 Samples of fingerprint images in modified sub-data sets

B (-70)	B (+35)	CW5	CW10	CCW5	CCW10
					

3.3 Minutiae-based Bit-strings CB

Minutiae-based bit-strings CB is proposed in [13] which the generation of the fingerprint template was performed by mapping the minutiae in the 3D array in order to remove the pre-alignment of the fingerprint requirement. The transformed fingerprint template is generated by mapping the minutiae in the 3D array and stored as bit-strings.

After scanning the finger and receiving the fingerprint image, the minutiae were extracted from the image and used as inputs in the template protection method. A 3D array

which consisted of cells was created and mapped the minutiae into the cells based on the position and the orientation of the reference minutiae. One of the minutiae was chosen as the reference minutia. The position of the reference minutia was at the centre of the 3D array. 1D bit-string was generated by sequentially visiting the cells in the 3D array by setting the cells in the 3D array. If there is more than one minutia in the cell, set '1' and otherwise '0'. Random permutation was applied on the bit-string using the user specific PIN.

Matching queried template with the enrolled template is performed by computing the similarity between them. Both local similarity and global similarity are computed. Local similarity is performed by comparing the values of the bit-strings of the query template with that of the enrolled template. Practically, two 1-D bit-strings generated from the same reference minutiae are the same. However, it cannot know which 1-D bit-strings are generated from the same reference point. Therefore, each bit-string of the query template is compared with that of the enrolled template. The similarity score between the enrolled bit-strings and the query bit-strings is computed by finding the number of common one between them with a bit-wise AND operator. In global similarity, the maximum of local similarity values is compared with the (decision) threshold. If the maximum local value is greater than or equal to the threshold, this local similarity value is picked for global similarity. Otherwise the value of the global similarity is set as '0'. The final similarity value is obtained. The sum of global similarity values is divided by the number of non-zero values of the global similarity. Finally, whether to accept or reject is decided by comparing the final similarity with the (decision) threshold.

3.4 Modified Minutiae-based Bit-strings CB

The process of the modified minutiae-based bit-strings CB fingerprint template generation is shown in Fig. 3. The idea is first introduced in [16]. The main idea is to use the similar process to generate the bit-string of the minutiae except the calculation of the reference point which is influent by [29]. Under the modified minutiae-based bit-strings CB, the minutiae are firstly extracted from the input fingerprint image. The template generation is performed by measuring the average distance of all the minutiae from the reference minutia and calculating the geometric information of the minutiae. One of the minutiae is selected as the reference minutia. For each of the reference minutia, the Euclidean distance is calculated with the other minutiae. Among these distances, the distance of the farthest minutia is selected. A line segment is drawn in the direction of the orientation of the reference minutia. The length of the line is the average distance. The horizontal direction is set as the reference direction. The slope of the line and its perpendicular is calculated. The transformed fingerprint templates are generated by checking the minutiae under the line or above the line. The generated template is permuted with the user pin which is the transformation key and used as the seed value to generate the cancellable fingerprint templates. The matching

process is done exactly as same as the minutiae-based bit-strings CB method.

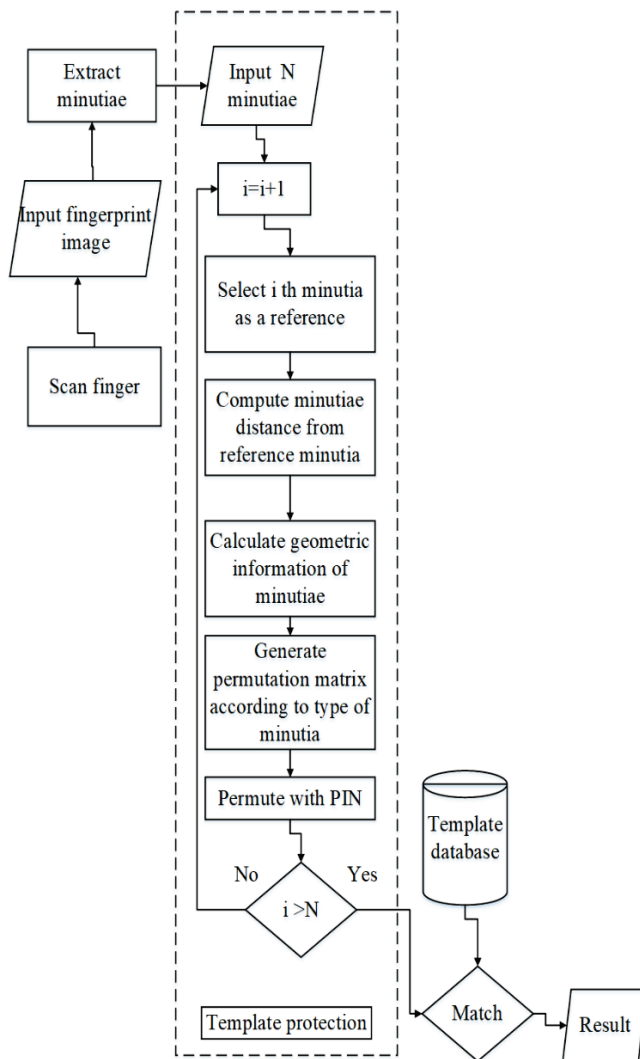


Figure 3 Modified minutiae-based bit-strings CB

3.5 Experimental Settings

The experiments are performed on Window 10, Intel (R) Core (TM) i7-8700U, 3.20 GHz processor and 8.00 GB RAM machine and implemented by using Matlab 2016a. The minutiae of every fingerprint are extracted and formatted according to the ISO standard by using the VeriFinger SDK 10.0 (trial version) of Neurotechnology [30]. The performance of the proposed method is evaluated by using the public fingerprint data set FVC2002DB1_B and FVC2002DB2_B [25] and their six modified data sets as described in the modified FVC database section. Tab. 3 shows the information of each data set used in this work.

For each fingerprint data set, three experiments are conducted. The first experiment aims to measure the accuracy of the same finger by matching the fingerprint impression and its modified images. The second experiment aims to measure the accuracy of the fingerprint impression against itself, its different fingerprint impressions and the

modified image of each fingerprint impressions. Thus, there are 640 comparisons for each data set. The third experiment aims to measure the accuracy of the fingerprint impression against only its different fingerprint impressions. Thus, there are 560 comparisons for each data set.

Table 3 Information of Each Data Set

FVC2002	Sensor Type	Image Size	Resolution
DB1_B	Optical Sensor	388×374	500 dpi
DB2_B	Optical Sensor	296×560	569 dpi

4 EXPERIMENTAL RESULTS AND DISCUSSIONS

Tab. 4 presents the accuracy performance of both methods on matching the same fingerprint image with its modified fingerprint image. Minutiae-based bit-strings CB is denoted CB1 while the modified minutiae-based bit-strings CB is denoted CB2. According to the results, the modified minutiae-based bit-strings CB (CB2) performs well on the FVC2002DB2_B data set. For the FVC2002DB1_B, minutiae-based bit-strings CB (CB1) performs poorly on the rotated images of 10 degrees both clockwise and anticlockwise direction.

Table 4 Accuracy of each method when comparing the same fingerprint image with its modified images

Image	FVC2002DB1_B		FVC2002DB2_B	
	CB1	CB2	CB1	CB2
B (-70)	100%	98.75%	100%	100%
B (+35)	98.75%	98.75%	100%	100%
CW5	100%	100%	87.50%	100%
CW10	86.25%	96.25%	71.25%	100%
CCW5	93.75%	98.75%	86.25%	100%
CCW10	87.50%	96.25%	73.75%	100%

Tab. 5 presents the accuracy performance of both methods on matching the fingerprint impression against itself, its different fingerprint impressions and the modified image of each fingerprint impressions. According to the results presented in Tab. 5, the modified minutiae-based bit-strings CB (CB2) outperforms the minutiae-based bit-strings CB (CB1) in all cases for both data sets. The modified minutiae-based bit-strings CB (CB2) also provides a higher accuracy on FVC2002DB2_B than that of FVC2002DB1_B. Interesting finding is that the modified minutiae-based bit-strings CB (CB2) seems to have an issue with a fingerprint impression of a dry finger B (+35) for both data sets.

Table 5 Accuracy of each method when comparing the fingerprint image with modified images of all impressions

Image	FVC2002DB1_B		FVC2002DB2_B	
	CB1	CB2	CB1	CB2
Origin	41.41%	96.56%	35%	100%
B (-70)	38.13%	96.41%	35.94%	100%
B (+35)	37.97%	95.63%	34.69%	99.84%
CW5	41.41%	96.56%	25.63%	100%
CW10	35.16%	96.41%	22.81%	100%
CCW5	35%	96.41%	25.31%	100%
CCW10	33.44%	95.31%	22.81%	100%

Tab. 6 presents the accuracy performance of both methods on matching the fingerprint image with different impressions of the same finger and the modified different

impressions of the same finger. As expected, the modified minutiae-based bit-strings CB (CB2) outperforms the minutiae-based bit-strings CB (CB1) in all cases for both data sets. The modified minutiae-based bit-strings CB (CB2) also provides a higher accuracy on FVC2002DB2_B than that of FVC2002DB1_B.

Table 6 Accuracy of each method when comparing the same fingerprint image with different impression of the same finger and its modified images

Image	FVC2002DB1_B		FVC2002DB2_B	
	CB1	CB2	CB1	CB2
Origin	28.91%	84.06%	25.71%	100%
B (-70)	25.63%	84.06%	26.79%	100%
B (+35)	28.91%	84.06%	25.36%	99.82%
CW5	24.38%	84.06%	16.79%	100%
CW10	24.38%	83.91%	15.89%	100%
CCW5	23.28%	84.06%	16.61%	100%
CCW10	22.50%	83.28%	15.54%	100%

From all experimental results, minutiae-based bit-strings CB (CB1) performs poorly on both data sets in comparison with that of the modified minutiae-based bit-strings CB (CB2). The fingerprint template generation in CB1 was dependent not only on the size of the image but also on the number of minutiae. The minutiae information in the modified images are different from that of the original images.

The modified minutiae-based bit-strings CB (CB2) is more robust when the image quality and the rotation is changed. This causes by the information collected by the CB2 in the minutiae because CB2 collects the distance and the geometric information of the minutia. Even the direction of the minutiae is changed, the Euclidean distance between the two minutiae remains the same. Moreover, the fingerprint images in FVC2002DB2_B are high resolution images. Therefore, the modified minutiae-based bit-strings CB (CB2) obtains a better matching accuracy on FVC2002DB2_B than that of FVC2002DB1_B.

Another interesting finding is that the minutiae-based bit-strings CB (CB1) results in a better performance on the modified images when the fingerprint image brightness is decreasing (i.e., B (-70)). This event is caused by the fact that most of the non-matched images of the origin images are collected in lighted or partial images. Fig. 4 shows the sample modified images in FVC2002DB2_B.

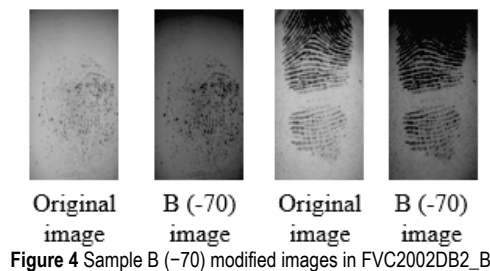


Figure 4 Sample B (-70) modified images in FVC2002DB2_B

According to the sample modified images, when the fingerprint image brightness is decreasing some minutiae points appear. Thus, more minutiae points can be extracted than that of the original images. The increasing number of minutiae points results in the increasing similarity values.

5 CONCLUSIONS

In the traditional fingerprint-based authentication system without any template protection, an unknown original fingerprint image can be reconstructed from the stored minutiae template. Obtaining minutiae templates is the loss of biometrics forever. If a few templates in the database are compromised, a re-enrolment of all the users may be necessary and will cost lot of time and money. Therefore, a protected fingerprint template has been used instead.

In the real situation, the quality of the fingerprint images may not be as expected for example during an emergency situation. Different qualities of fingerprints are created according to environmental conditions during an emergency, such as wet or dry fingers and a disoriented angle of fingerprint images. These fingerprint images are used for evaluating two cancellable fingerprint template approaches. The two cancellable fingerprint template approaches have been selected because their techniques include the geometric information of the fingerprint into the minutiae. As a result, the two approaches have a high potential to be robust to the disoriented angle of the fingerprint images.

The experimental results have shown that the modified minutiae-based bit-strings CB performs well on all conditions during an emergency situation by achieving the matching accuracy between 83% and 100% on FVC2002DB1_B data set and between 99% and 100% on FVC2002DB2_B data set. The result of this work will be further used for proposing an improved fingerprint feature transformation approach for personal health record users during emergency situation as introduced in [31] under the framework proposed in [32].

Acknowledgment

This work was supported by the Higher Education Research Promotion and the Thailand's Education Hub for Southern Region of ASEAN Countries Project Office of the Higher Education Commission (Grant number THE-AC 033/2016).

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

6 REFERENCES

- [1] Maltoni, D., Maio, D., Jain, A. K., & Prabhakar, S. (2009). *Handbook of fingerprint recognition* (2nd ed.). Springer-Verlag. <https://doi.org/10.1007/978-1-84882-254-2>
- [2] Jain, A. K., Nandakumar, K., & Ross, A. (2016). 50 years of biometric research: accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79, 80-105. <https://doi.org/10.1016/j.patrec.2015.12.013>
- [3] Ratha, N. K., Chikkerur, S., Connell, J. H., & Bolle, R. M. (2007). Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4), 561-572. <https://doi.org/10.1109/TPAMI.2007.1004>

- [4] Cao, K., & Jain, A. K. (2014). Learning fingerprint reconstruction: from minutiae to image. *IEEE Transactions on Information Forensics and Security*, 10(1), 104-117. <https://doi.org/10.1109/TIFS.2014.2363951>
- [5] Jin, Z., Hwang, J. Y., Lai, Y. L., Kim, S., & Teoh, A. B. J. (2018). Ranking-based locality sensitive hashing-enabled cancelable biometrics: index-of-max hashing. *IEEE Transactions on Information Forensics and Security*, 13(2), 393-407. <https://doi.org/10.1109/TIFS.2017.2753172>
- [6] Wang, S., Deng, G., & Hu, J. (2017). A partial Hadamard transform approach to the design of cancelable fingerprint templates containing binary biometric representations. *Pattern Recognition*, 61, 447-458. <https://doi.org/10.1016/j.patcog.2016.08.017>
- [7] Wang, S., Yang, W., & J. Hu. (2017). Design of alignment-free cancelable fingerprint templates with zoned minutia pairs. *Pattern Recognition*, 66, 295-301. <https://doi.org/10.1016/j.patcog.2017.01.019>
- [8] Ferrara, M., Maltoni, D., & Cappelli, R. (2012). Noninvertible minutia cylinder-code representation. *IEEE Transactions on Information Forensics and Security*, 7(6), 727-737. <https://doi.org/10.1109/TIFS.2012.2215326>
- [9] Cappelli, R., Ferrara, M., & Maltoni, D. (2010). Minutia cylinder-code: a new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12), 2128-2141. <https://doi.org/10.1109/TPAMI.2010.52>
- [10] Ferrara, M., Maltoni, D., & Cappelli, R. (2014). A two-factor protection scheme for MCC fingerprint templates. In *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1-8. IEEE.
- [11] Wang, S. & Hu, J. (2016). A blind system identification approach to cancelable fingerprint templates. *Pattern Recognition*, 54, 14-22. <https://doi.org/10.1016/j.patcog.2016.01.001>
- [12] Prasad, M. V., Anugu, J. R., & Rao, C. R. (2016). Fingerprint template protection using multiple spiral curves. In *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, Springer, New Delhi, 593-601. https://doi.org/10.1007/978-81-322-2538-6_61
- [13] Lee, C. & Kim, J. (2010). Cancelable fingerprint templates using minutiae-based bit-strings. *Journal of Network and Computer Applications*, 33(3), 236-246. <https://doi.org/10.1016/j.jnca.2009.12.011>
- [14] Guo, L., Mao, Y., & Guo, Y. (2016). Non-invertible fingerprint template protection with polar transformations. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, 730-735, IEEE. <https://doi.org/10.1109/PST.2016.7906990>
- [15] Das, P., Karthik, K., & Garai, B. C. (2012). A robust alignment-free fingerprint hashing algorithm based on minimum distance graphs. *Pattern Recognition*, 45(9), 3373-3388. <https://doi.org/10.1016/j.patcog.2012.02.022>
- [16] Kumar, N. (2012). *Cancelable fingerprint template*. Department of Mathematics, Indian Academy of Sciences.
- [17] Wong, W. J., Wong, M. L. D., & Kho, Y. H. (2013). Multi-line code: A low complexity revocable fingerprint template for cancelable biometrics. *Journal of Central South University*, 20(5), 1292-1297. <https://doi.org/10.1007/s11771-013-1614-8>
- [18] Belguechi, R., Hafiane, A., Cherrier, E., & Rosenberger, C. (2016). Comparative study on texture features for fingerprint recognition: application to the biohashing template protection scheme. *Journal of Electronic Imaging*, 25(1), 013033. <https://doi.org/10.1117/1.JEI.25.1.013033>
- [19] Moujahdi, C., Bebis, G., Ghouzali, S., & Rziza, M. (2014). Fingerprint shell: Secure representation of fingerprint template. *Pattern Recognition Letters*, 45, 189-196. <https://doi.org/10.1016/j.patrec.2014.04.001>
- [20] Nazmul, R., Islam, M. R., & Chowdhury, A. R. (2017). Alignment-Free Fingerprint Template Protection Technique Based on Minutiae Neighbourhood Information. In *International Conference on Applications and Techniques in Cyber Security and Intelligence*, 256-265, Edizioni della Normale, Cham. https://doi.org/10.1007/978-3-319-67071-3_32
- [21] Sandhya, M. & Prasad, M. V. (2015). k-Nearest Neighborhood Structure (k-NNS) based alignment-free method for fingerprint template protection. In *2015 International Conference on Biometrics (ICB)*, 386-393, IEEE. <https://doi.org/10.1109/ICB.2015.7139100>
- [22] Li, G., Yang, B., Rathgeb, C., & Busch, C. (2015). Towards generating protected fingerprint templates based on bloom filters. In *3rd International workshop on biometrics and forensics (IWBF 2015)*, 1-6, IEEE. <https://doi.org/10.1109/IWBF.2015.7110224>
- [23] Kanagalakshmi, K. & Chandra, E. (2013). A Novel Technique for Cancelable and Irrevocable Biometric Template Generation for Fingerprints. *Global Journal of Computer Science and Technology*, 13(6). Retrieved from <https://computerresearch.org/index.php/computer/article/view/208/208>
- [24] See <http://bias.csr.unibo.it/fvc2000/databases.asp>.
- [25] See <http://bias.csr.unibo.it/fvc2002/databases.asp>.
- [26] See <http://bias.csr.unibo.it/fvc2004/databases.asp>.
- [27] See <http://bias.csr.unibo.it/fvc2006/databases.asp>.
- [28] Choosang, P. (2016). Secure fingerprint identification evaluation for unconscious personal health records users: a case study in an emergency situation, *Master's thesis*, Prince of Songkla University, Thailand.
- [29] Ang, R., Safavi-Naini, R., & McAven, L. (2005). Cancelable key-based fingerprint templates. In *Australasian Conference on Information Security and Privacy*, Springer, Berlin, Heidelberg, 242-252. https://doi.org/10.1007/11506157_21
- [30] See <http://neurotechnology.com>.
- [31] Choosang, P. & Vasupongayya, S. (2015). Using fingerprints to identify personal health record users in an emergency situation. In *2015 International Computer Science and Engineering Conference (ICSEC)*, 1-6, IEEE. <https://doi.org/10.1109/ICSEC.2015.7401421>
- [32] Thummavet, P. & Vasupongayya, S. (2015). Privacy-preserving emergency access control for personal health records. *Maejo International Journal of Science and Technology*, 9(1), 108-120. Retrieved from <http://www.mijst.mju.ac.th/vol9/108-120.pdf>

Authors' contacts:

Ei Ei MON, M.C.Sc*
5910130005@psu.ac.th

Sangsuree VASUPONGAYYA, PhD, Assist. Prof.
vsangsur@coe.psu.ac.th

Montri KARNJANADECHA, PhD, Assoc. Prof.
montri@coe.psu.ac.th

Touchai ANGCHUAN, MSc, Assist. Prof.
touch@coe.psu.ac.th

Department of Computer Engineering,
Faculty of Engineering, Prince of Songkla University,
Hat Yai, Songkhla 90112, Thailand

MULTIPLICATION OF MEDIUM-DENSITY MATRICES USING TENSORFLOW ON MULTICORE CPUs

Siraphob THEERACHEEP, Jaruloj CHONGSTITVATANA

Abstract: Matrix multiplication is an essential part of many applications, such as linear algebra, image processing and machine learning. One platform used in such applications is TensorFlow, which is a machine learning library whose structure is based on dataflow programming paradigm. In this work, a method for multiplication of medium-density matrices on multicore CPUs using TensorFlow platform is proposed. This method, called *tbt_matmul*, utilizes TensorFlow built-in methods *tf.matmul* and *tf.sparse_matmul*. By partitioning each input matrix into four smaller sub-matrices, called tiles, and applying an appropriate multiplication method to each pair depending on their density, the proposed method outperforms the built-in methods for matrices of medium density and matrices of significantly uneven distribution of non-zeros.

Keywords: Sparse matrix; Matrix multiplication; TensorFlow

1 INTRODUCTION

Matrix multiplication is a basis for computation in many areas, such as linear algebra, machine learning and image processing. Various implementations of matrix multiplication are studied in different environments. For dense matrices, many efficient matrix multiplication algorithms such as Strassen algorithm [1] and Coppersmith-Winograd algorithm [2] are proposed. In these algorithms all matrices are treated as dense matrices. Another approach to improve the efficiency of matrix multiplication is based on the sparsity of matrices. In [3], a hardware accelerator for neural networks, utilizing irregularity in sparse neural networks and the sparsity of weight matrices, is developed. In [4], a sparse tensor representation, F-COO, is proposed, and, based on F-COO, sparse tensors operations and sparse matrix multiplication on GPU are optimized.

TensorFlow [5] also provides the matrix multiplications for both dense matrices and sparse matrices. Like many other lower-level functions, these matrix multiplication methods are built on various kernel-implementations, depending on the platform. For example, methods in Eigen [6] are built for dense and sparse matrix multiplication on CPUs and methods in cuBLAS [7] are built for dense matrix multiplication on GPUs.

Medium-density matrices cannot benefit from matrix multiplication designed for sparse matrices. However, a medium-density matrix can be divided into sub-matrices, some of which are dense, and some are sparse. In this paper, we propose an approach to reduce the computation time for medium-density matrix multiplication. This approach divides a medium-density matrix into four equal-size sub-matrices and chooses an appropriate matrix multiplication method for each pair of sub-matrices, based on the density of the sub-matrices.

Using this approach, we implement a matrix multiplication function for medium-density matrices on TensorFlow. An experiment is performed to compare the proposed method to TensorFlow's matrix multiplication for

dense matrices and that for sparse matrices. It is found that the proposed method is faster than both matrix multiplication methods in TensorFlow for medium-density matrices.

The proposed approach can be applied for matrix multiplication on any platform, and any efficient implementation of matrix multiplication can be used for sub-matrices. This is an advantage of this approach because it can benefit from any performance improvement in the low-level libraries.

The remaining sections of this paper are organized as follows. Section 2 presents related works on matrix multiplication for tensors, dense matrices and sparse matrices on various environments to motivate the need for efficient multiplication for medium density matrices. Section 3 describes the proposed matrix multiplication algorithm for medium-density matrices. The performance evaluation of the proposed method is discussed in Section 4. Section 5 presents conclusion and future works.

2 BACKGROUND AND MOTIVATION

Since matrix multiplication is often used in many applications, various implementations of matrix multiplication are studied in different environments. A hardware accelerator for neural network [3] which also includes matrix multiplication, was proposed. The accelerator utilized the sparsity and irregularity of weight matrices of neural network models to improve the computation efficiency. Many efficient matrix multiplication algorithms are proposed for different architectures. For example, in [8], a table and heap-based algorithm for sparse matrix multiplication is optimized for multicore and Intel KNL processors. This implementation outperforms Intel MKL sparse matrix multiplication for large matrices. In [9], software libraries for linear algebra computation, including matrix multiplication, are developed for GPUs and CPUs. Nvidia cuBLAS library [7] is a linear algebra library for Nvidia's GPU. Intel MKL [10] is a mathematical library optimized for Intel's processor. OpenBLAS [11] is an open-

source linear algebra library. Higher-level libraries, such as Numpy, scikit-learn, Theano, PyTorch and TensorFlow, are built on these linear algebra and mathematical libraries in order to utilize performance optimization of low-level libraries.

Many of these implementations are designed for dense matrices. On multicore CPUs, dense matrix multiplication can be made faster by maximizing the parallelism while optimizing the overhead. When the operations are performed on sparse matrices, many operations are wasted on multiplying zeroes and no performance improvement can be gained from the sparsity of matrices. The performance gain for sparse matrix multiplication depends on the representation of sparse matrices. F-COO [4] is proposed as a representation of sparse tensors which is used for tensor operations on GPU. A sparse tensor-times- dense matrix multiplication (SpTTM) [12] is implemented for both CPU and GPU platforms.

TensorFlow, developed by Google, is an open-source machine learning library that contains various methods that help in the development and training of machine learning models. It also includes low level functions such as matrix multiplication. On GPUs, TensorFlow provides the matrix multiplication only for dense matrices. But, on CPUs, it provides the matrix multiplication for both dense matrices and sparse matrices. Since TensorFlow provides kernel implementations for many platforms, these multiplication functions are optimized for different platforms. It is suggested that, using TensorFlow on multicore CPUs, the multiplication method for sparse matrices could be used, instead of the dense matrix multiplication method, when the density of the input matrices is lower than 70% [13].

However, a matrix can be divided into some dense sub-matrices and some sparse sub-matrices, and the multiplication of a pair of matrices can be decomposed into the multiplication of pairs of sub-matrices. The idea of multiplying sub-matrices is called tiled matrix multiplication, which is used to perform parallel matrix multiplication by partitioning a pair of matrices into multiple ‘tiles’ and distribute them to different computing units to perform concurrently. This method allows multiple parts of the result matrix to be computed simultaneously [14], [15], [16]. Based on tiled matrix multiplication, an efficient multiplication method can be chosen for each pair of tiles, instead of using one multiplication method for the whole matrix.

3 PROPOSED MULTIPLICATION FOR MEDIUM-DENSITY MATRICES

We propose an approach for matrix multiplication which is optimized for matrices with medium density. Each of the two input matrices is divided into four sub-matrices. An $m \times n$ matrix is divided into four $m/2 \times n/2$ sub-matrices. Pairs of sub-matrices are multiplied, using an appropriate matrix multiplication method. In order to optimize the computation time, an appropriate multiplication method is chosen for each pair of sub-matrices based on their density. The proposed method, referred to as two-by-two matrix multiplication function *tbt_matmul*, is described below.

Function: *tbt_matmul(A, B, C)*

Let A , B and C be matrices of size $(m \times n)$, $(n \times p)$ and $(m \times p)$, respectively.

Algorithm:

1. Divide A into A_{11} , A_{12} , A_{21} , and A_{22} of size $m/2 \times n/2$.
2. Divide B into B_{11} , B_{12} , B_{21} , and B_{22} of size $n/2 \times p/2$.
3. Find the density of A_{11} , A_{12} , A_{21} , and A_{22} as D_{A11} , D_{A12} , D_{A21} and D_{A22} , respectively.
4. Find the density of B_{11} , B_{12} , B_{21} , and B_{22} as D_{B11} , D_{B12} , D_{B21} and D_{B22} , respectively.
5. Choose the matrix multiplication method for each pair of sub-matrices.

$$A_{11}B_{11}matmulfn = getMatmulFn(D_{A11}, D_{B11})$$

$$A_{12}B_{21}matmulfn = getMatmulFn(D_{A12}, D_{B21})$$

$$\dots$$

$$A_{22}B_{22}matmulfn = getMatmulFn(D_{A22}, D_{B22})$$
6. Use the chosen multiplication method to multiply each pair of sub-matrices.

$$C_{A11B11} = A_{11}B_{11}matmulfn(A_{11}, B_{11})$$

$$C_{A12B21} = A_{12}B_{21}matmulfn(A_{12}, B_{21})$$

$$\dots$$

$$C_{A22B22} = A_{22}B_{22}matmulfn(A_{22}, B_{22})$$
7. Find sub-matrices of the result matrix C from matrices obtained in Step 6.

$$C_{11} = C_{A11B11} + C_{A12B21}$$

$$C_{12} = C_{A11B12} + C_{A12B22}$$

$$C_{21} = C_{A21B11} + C_{A22B21}$$

$$C_{22} = C_{A21B12} + C_{A22B22}$$
8. Combine sub-matrices C_{11} , C_{12} , C_{21} and C_{22} into C .

This algorithm is implemented on Python 3.6.5, TensorFlow API r1.11 using TensorFlow’s matrix multiplications, *tf.matmul* and *tf.sparse_matmul*. The method *tf.matmul* is used when both operands are dense matrices, while the method *tf.sparse_matmul* is used when at least one of the operands is a sparse matrix. For the method *tf.sparse_matmul(A, B, A_is_sparse, B_is_sparse)*, the parameters A and B are the input matrices, and the parameters *A_is_sparse* and *B_is_sparse* are boolean values specifying if the input matrices are sparse.

Table 1 Matrix Multiplication Method Chosen for Sub-matrices

Density of B	Density of A				
	0-20	20-30	30-40	40-50	50-100
0-10	$tf.sparse_matmul(A, B,$ $A_is_sparse=False,$ $B_is_sparse=True)$				
10-20					
20-30					
30-40	$tf.sparse_matmul(A, B,$ $A_is_sparse=True,$ $B_is_sparse=True)$				$tf.matmul(A, B)$
40-100	$tf.sparse_matmul(A, B,$ $A_is_sparse=True, B_is_sparse=False)$				

The criteria for choosing between *tf.matmul* and *tf.sparse_matmul* is based on the multiplication time for varying density of input matrices on the platform used to evaluate the performance. For the multiplication of sub-matrices, if a matrix is sparse enough that *tf.sparse_matmul* is faster than *tf.matmul*, *tf.sparse_matmul* is chosen.

Otherwise, *tf.matmul* is used. Using different multiplication methods for each multiplication of sub-matrices can reduce the computation time if the performance difference of the *tf.matmul* and *tf.sparse_matmul* outweighs the overhead cost of partitioning and merging matrices. The criteria for choosing the multiplication method for sub-matrices shown in Tab. 1 is used in the function *getMatmulFn* in this paper.

4 PERFORMANCE EVALUATION

tbt_matmul is our implementation of the proposed matrix multiplication method on Python 3.6.5 with TensorFlow API r1.11. In this implementation, *tf.matmul* is used for dense matrices and *tf.sparse_matmul* is used for sparse matrices. *tbt_matmul* is compared with the multiplication using *tf.matmul* and *tf.sparse_matmul* for the whole matrices. The experiments were performed on a Laptop with Intel Core i7-4720HQ quad-core CPU, 16GB DDR3 RAM. The operating system used in this experiment was Linux Ubuntu 18.04.2 with Python 3.6.5 and TensorFlow API r1.11. In the experiments, the default configuration for thread pool was used, and the number of threads is set to maximum [17]. Thus, TensorFlow used eight threads in this experiment.

We performed the experiment on matrices with various overall density, ranging from 0.3 to 0.8. Furthermore, the performance of the proposed method is evaluated on matrices with different combination of higher-density and lower-density sub-matrices. These matrices are randomly generated so that two of their sub-matrices have higher density of non-zeros and two have lower density of non-zeros as shown in Fig. 1. The non-zero distribution of sub-matrices is varied, as shown in Tab. 2. Twenty pairs of matrices are generated for each non-zero matrices, and the average computation time is used for the comparison.

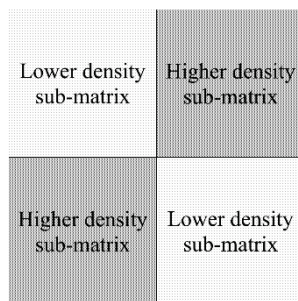


Figure 1 The density distribution pattern of input matrices

Table 2 Non-zero Distribution of Sub-matrices

Overall Density	(Higher Density, Lower Density)			
0.3	(0.5, 0.1)	(0.4, 0.2)		
0.4	(0.7, 0.1)	(0.6, 0.2)	(0.5, 0.3)	
0.5	(0.9, 0.1)	(0.8, 0.2)	(0.7, 0.3)	(0.6, 0.4)
0.6	(1.0, 0.2)	(0.9, 0.3)	(0.8, 0.4)	(0.7, 0.5)
0.7	(1.0, 0.4)	(0.9, 0.5)	(0.8, 0.6)	
0.8	(1.0, 0.6)	(0.9, 0.7)		

In order to evaluate the proposed method for different sizes of input, the experiments were performed on large

matrices, i.e. 12,000×12,000 matrices, and small matrices, i.e. 3,000×3,000 matrices.

The graphs in Figs. 2-5 show the execution time for all three methods for matrices with different densities and sizes. The execution time for all three methods is grouped together for the comparison. The x-axis shows the density and non-zero distribution of the input matrices. The non-zero distribution of matrices is labelled in the graphs as (*d*, *s*), where *d* is the density of the higher-density sub-matrices and *s* is the density of the lower-density sub-matrices. The overall density is the average of *d* and *s*. For example, (0.5, 0.1) indicates that the density of the higher-density sub-matrices is 0.5, the density of lower-density sub-matrices is 0.1, and overall density is 0.3.

In section 4.1-4.3, the performance of the three methods are compared on large matrices of the size 12,000×12,000. Section 4.1 shows the performance on medium-density matrices, section 4.2 shows the performance on low-density matrices, and section 4.3 shows the performance on high-density matrices. Furthermore, section 4.4 compares the performance of the three methods on small, medium-density matrices of the size 3,000×3,000.

4.1 On Medium-density Matrices

Fig. 2 shows the execution time of the three multiplication methods on medium-density matrices of 50% and 60% density. The proposed method performs better than *tf_sparse_matmul* and *tf.matmul* for almost all non-zero distribution. When the density difference between the higher-density sub-matrices and the lower-density sub-matrices is small, i.e. for the non-zero distribution of (0.6, 0.4) and (0.7, 0.5), the proposed method is not the fastest. However, it is faster than *tf_sparse_matmul*, but slower than *tf.matmul* for the non-zero distribution of (0.6, 0.4).

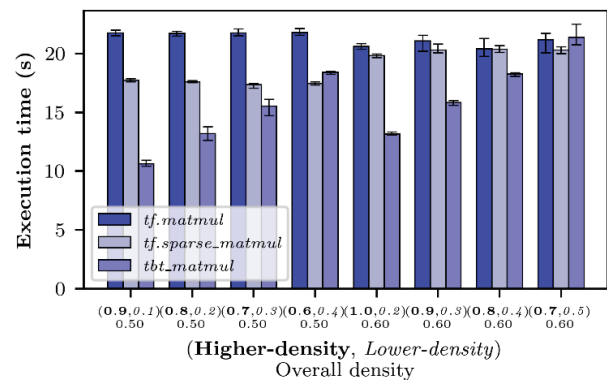


Figure 2 Execution time of *tf.matmul*, *tf.sparse_matmul* and *tbt_matmul* for 12000×12000 matrices with 50-60% density.

Furthermore, all three methods took almost the same amount of time for the non-zero distribution of (0.7, 0.3).

4.2 On Low-density Matrices

The execution time of the three multiplication methods on low-density matrices of 30% and 40% density is shown in Fig. 3. As suggested in [13], the method *tf.matmul* is not

suitable for sparse matrices, and takes longest time of the three. The proposed method performs better than *tf.sparse_matmul* when the density of the higher-density and lower-density sub-matrices are significantly different, i.e. the non-zero distribution of **(0.5, 0.1)**, **(0.7, 0.1)** and **(0.6, 0.2)**. But when the density difference between the higher-density and the lower-density sub-matrices are small, our method is minutely slower than *tf.sparse_matmul*.

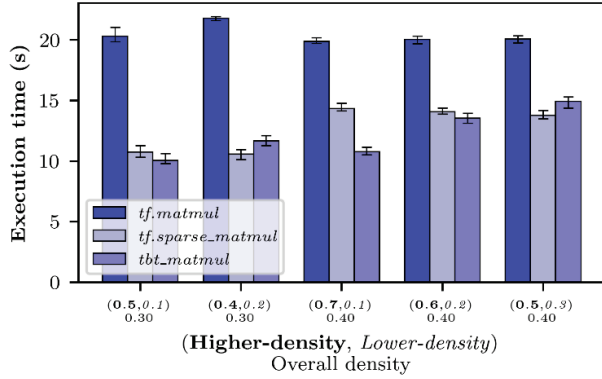


Figure 3 Execution time of *tf.matmul*, *tf.sparse_matmul* and *tbt_matmul* for 12000×12000 matrices with 30-40% density.

4.3 On High-density Matrices

Fig. 4 shows the execution time of the three multiplication methods on high-density matrices of 70% and 80% density. As expected, the method *tf.sparse_matmul* is not good for dense matrices and takes longest time of the three. For matrices with 70% density, the proposed method performs better than, or equally well as *tf.matmul* when the density of the higher-density and lower-density sub-matrices are significantly different, i.e. **(1.0, 0.4)** and **(0.9, 0.5)**.

For matrices with 80% density, our proposed method is slower than *tf.matmul*, regardless of the non-zero distribution.

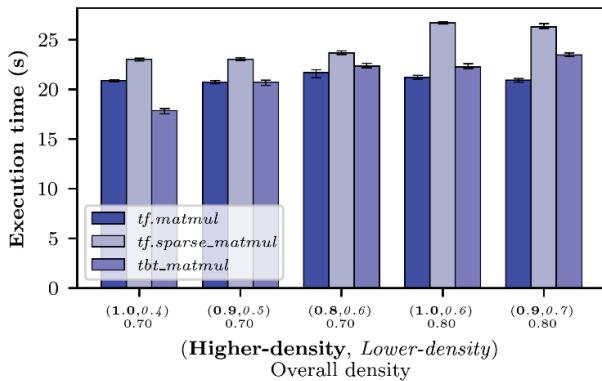


Figure 4 Execution time of *tf.matmul*, *tf.sparse_matmul* and *tbt_matmul* for 12000×12000 matrices with 70-80% density.

4.4 On Small Matrices with Medium-density

For input matrices of size 3,000×3,000, *tf.matmul* performs best on matrices with density at least 70% but was outperformed by *tf.sparse_matmul* on matrices with density under 70%. On the other hand, the method *tbt_matmul* is the

slowest in all datasets most likely due to overhead of matrix partitioning and merging outweighing the performance gained. The execution time on small medium-density matrices is shown in Fig. 5

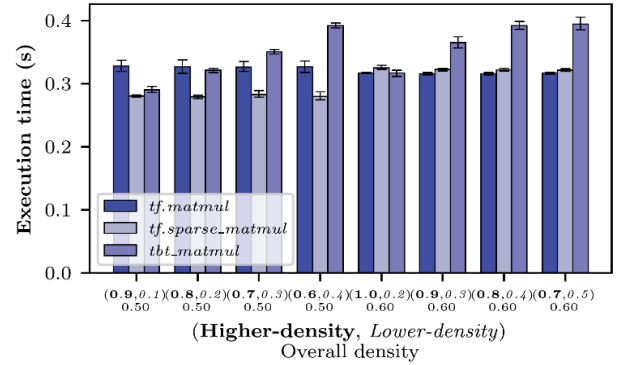


Figure 5 Execution time of *tf.matmul*, *tf.sparse_matmul* and *tbt_matmul* for 3000×3000 matrices with 50-60% density.

Based on the experiment, the density difference between the higher-density and lower-density sub-matrices is the important factor for the performance of our proposed method. When density difference between the higher-density and lower-density sub-matrices is large, the method *tbt_matmul* outperforms the other two methods for the input matrices with density of 70% or lower. However, for medium-density matrices, the proposed method is faster or comparable to the other two for all non-zero distribution. For sparse matrices, the proposed method is comparable to the *tf.sparse_matmul*. For dense matrices, the proposed method is comparable to *tf.matmul*. Furthermore, our proposed method is not suitable for small matrices.

5 DISCUSSION

The computation time of *tf.matmul* does not depend on the density of input matrices because it treats all matrices regardless of their density. It only performs better than the other two methods when the density of input matrices are very high.

The method *tf.sparse_matmul* takes into account the density of the input matrices and therefore performs better than *tf.matmul* at density lower than 70%. However, its performance is the same for different distribution of non-zeros in the matrices.

On the other hand, the performance of the proposed method *tbt_matmul* depends on both the matrix density and the density difference between the higher-density and lower-density sub-matrices of the input matrices. For medium density matrices, *tbt_matmul* is the fastest, especially when the density difference between the higher-density and lower-density sub-matrices is large. The performance gain can be attributed to the selection of the appropriate multiplication methods in step 5 of the algorithm. For example, for the input dataset **(0.9, 0.1)**, 6 out of 8 multiplications of sub-matrices use *tf.sparse_matmul* and two multiplication of sub-matrices use *tf.matmul*. Using dense matrix multiplication for these two pairs of sub-matrices can reduce the computation time,

compared to using sparse matrix multiplication for the whole matrix. On the other hand, when the difference between the higher-density and lower-density sub-matrices is small, such as (1.0, 0.6) or (0.7, 0.5), the multiplication method chosen for all pairs of sub-matrices pair is *tf.matmul* in many cases, because the density of each submatrix is relatively high. As a result, we do not get the performance gain because it is the same as using that multiplication method for the whole matrix, with additional cost of splitting and merging.

In this work, a matrix is divided into 4 sub-matrices. Further dividing a matrix into smaller sub-matrices, such as 16, does not lead to better performance, most likely due to the size of the dataflow graphs required for the computation.

6 CONCLUSION

In this work, we have proposed a multiplication method, *tbt_matmul*, for medium-density matrices on TensorFlow. The proposed method divides each input matrix into four sub-matrices and multiplies them using one of the TensorFlow built-in matrix multiplication methods, *tf.matmul* or *tf.sparse_matmul*, depending on the density of each sub-matrices. The method outperforms *tf.matmul* when the input matrices have overall density of 70% or lower and outperforms *tf.sparse_matmul* when the input matrices have overall density of 40% or higher. This method performs especially well when the densities of sub-matrices are significantly different.

The concept of the proposed method, multiplying different sub-matrices with the appropriate multiplication methods, can be used in any platform with efficient multiplication methods for dense matrices and for sparse matrices. It can benefit from the matrix multiplication methods, which are optimized on a specific platform.

For the future work, this concept can also be applied for matrix multiplication on GPUs. However, no sparse matrix multiplication is provided for GPUs on TensorFlow and it should be implemented in order to further improve matrix multiplication on GPU.

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

7 REFERENCES

- [1] Volker, S. (1969). Gaussian elimination is not optimal. *Numerische Mathematik*, 13, 354-356. <https://doi.org/10.1007/BF02165411>
- [2] Don, C. & Shmuel, W. (1990). Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, 251-280. [https://doi.org/10.1016/S0747-7171\(08\)80013-2](https://doi.org/10.1016/S0747-7171(08)80013-2)
- [3] Shijin Z., Zidong D., Lei Z., Huiying L., Shaoli L., Ling L., Qi G., Tianshi, C., & Yunji, C. (2016). Cambricon-X: An accelerator for sparse neural networks. *49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Taipei, 1-12.

- [4] Bangtian, L., Chengyao, W., Anand, D. S., & Maryam, M. D. (2017). A unified optimization approach for sparse tensor operations on GPUs. *2017 IEEE International Conference on Cluster Computing (CLUSTER)*, Honolulu, HI, 47-57.
- [5] Abadi, M., Barham, P., Chen, J. et al. (2016). TensorFlow: a system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*.
- [6] Benoît, J. & Gaël, G. Eigen library for linear algebra. <http://eigen.tuxfamily.org/>
- [7] NVIDIA Corporation. (2008). Santa Clara, California, Cublas library, <https://developer.nvidia.com/cublas>
- [8] Yusuke, N., Satoshi, M., Ariful, A., & Aydin, B. (2018). High-performance sparse matrix-matrix products on Intel KNL and multicore architectures. *ICPP '18*.
- [9] Gudula, R. & Michael, S. (2010). Fast recursive matrix multiplication for multi-core architectures. *ICCS 2010*, 67-76. <https://doi.org/10.1016/j.procs.2010.04.009>
- [10] Intel Corporation, Intel® Math Kernel Library, <https://software.intel.com/en-us/mkl>
- [11] Zhang, X., Martin, K., Werner, S., Wang, Q., Zaheer, C., Chen, S., & Luo, W. *OpenBlas*. <https://www.openblas.net>
- [12] Jiajia, L., Yuchen, M., Chenggang, Y., & Richard, V. (2016). Optimizing sparse tensor times matrix on multi-core and many-core architectures. *6th Workshop on Irregular Applications: Architecture and Algorithms (IA3)*, Salt Lake City, UT, 26-33.
- [13] TensorFlow Python API Documentation, https://www.tensorflow.org/api_docs/python.
- [14] Chien, S. W. D., Markidis, S., Olshevsky, V., Bulatov, Y., Laure, E., & Vetter, J. S. (2019). TensorFlow doing HPC: an evaluation of TensorFlow performance in HPC applications. *The Ninth International Workshop on Accelerators and Hybrid Exascale Systems (AsHES'19)*, Rio de Janeiro, Copacabana, Brazil. <https://doi.org/10.1109/IPDPSW.2019.00092>
- [15] Minwoo, K., Yong, J. J., & Won, W. R. (2011). Parallel transpose of matrix multiplication based on the tiling algorithms. *IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Seoul, 1-3.
- [16] Haitao, W., Guang, R. G., & Elkin, G. (2015). Energy efficient multi-level tiling for dense matrix multiplication on many-core architecture. *Sixth International Green and Sustainable Computing Conference (IGSC)*, Las Vegas, NV, 1-6. <https://doi.org/10.1109/IGCC.2015.7393735>
- [17] TensorFlow Performance Optimization Guide, <https://www.tensorflow.org/guide/performance/overview>

Authors' contacts:

Siraphob THEERACHEEP
Department of Mathematics and Computer Science,
Faculty of Science, Chulalongkorn University,
Bangkok 10330, Thailand
siraphob47@gmail.com

Jaruloj CHONGSTITVATANA, Assistant Professor
Department of Mathematics and Computer Science,
Faculty of Science, Chulalongkorn University,
Bangkok 10330, Thailand
jaruloj@gmail.com

CONTRAST AND COLOR BALANCE ENHANCEMENT FOR NON-UNIFORM ILLUMINATION RETINAL IMAGES

Preecha VONGHIRANDECHA, Montri KARNJANADECHA, Sathit INTAJAG

Abstract: Color retinal images play an important role in supporting a medical diagnosis. However, some retinal images are unsuitable for diagnosis due to the non-uniform illumination. In order to solve this problem, we propose a method for improving non-uniform illumination that can enhance the image quality of a color fundus photograph suitable for reliable visual diagnosis. Firstly, a hidden anatomical structure in dark regions of the retinal images is revealed by improving the image luminosity with gamma correction. Secondly, multi-scale tone manipulation is then used to adjust the image contrast in the lightness channel of $L^*a^*b^*$ color space. Finally, color balance is adjusted by specifying the image brightness based on Hubbard's specification. The performance of the applied method has been evaluated against the data from the DIARETDB1 dataset. The results obtained show that the proposed algorithm performs well for correcting the non-uniform illumination of color retinal images.

Keywords: human visual perception; image contrast and color balance enhancement; non-uniform illumination; retinal image

1 INTRODUCTION

According to the World Health Organization, eye diseases such as age-related macular degeneration (AMD), cataract, and glaucoma are the main cause of blindness in the elderly worldwide [1, 2]. Most current diagnostic systems are based on color retinal photography. However, the image may be unsatisfactory for diagnosis due to their low quality caused by camera properties, non-uniform illumination, and the experience of photographers [3]. Consequently, image enhancement algorithms are widely used to improve the image quality to provide better visibility of retinal anatomical structures.

Histogram equalization (HE) is a common contrast enhancement technique which utilizes the cumulative distribution function (CDF) of the input image for stretching the dynamic range of intensity levels. The main drawback of HE is that it tends to shift the mean intensity value to the middle of the dynamic range. In some cases, HE provides the results over enhancement or unnatural appearance. To overcome the mentioned drawback, various methods including BPHEME [4], BPDHE [5], and BPDFHE [6] are proposed to preserve image brightness. However, those algorithms are not appropriate when the input image is underexposed or overexposed because they provide output images with a mean brightness close to the input and some results in a manner unsuitable for human perception.

Recently, image enhancement based on luminosity and contrast adjustment (LCA) [7], which augments HE, has been designed for a color retinal image. The first part of LCA creates a luminance gain matrix by correcting the gamma value of the value channel, V , in the hue-saturation-value (HSV) color space, and it uses the gain value to enhance the intensity of R , G , and B channels. In the second part, the contrast is improved by converting the enhanced image to the $L^*a^*b^*$ color space and applying CLAHE [8] to adjust the local contrast of the luminance L with the number of tiles and clip

limit equal to 8×8 and 0.01, respectively. LCA could handle the contrast enhancement, but neglects the color balance.

In 2016, Dai et al. [9] proposed retinal fundus image enhancement by using normalized convolution and noise removal (NCNR). The original image is used to extract the information in the background image by employing a normalized convolution filter [10]. The difference between the original image and the background image is then multiplied by a contrast factor and recombined with the original image in order to increase the image contrast. Finally, the fourth order PDEs [11] and a relaxed median filter [12] are applied to reduce noise in the image. Although this technique enhances detailed information by increasing the image contrast, especially in the vessel areas, the color balance of the output image is not proper for human perception and the NCNR method is time consuming.

The contrast enhancement methods, as described above, usually provide results in unnatural looks and displeasing color balance. Motivated and inspired by the findings published in [13], Hubbard et al. analyzed brightness, contrast, and color balance of digital images compared to film retinal images in the AREDS reading center for the AMD screening. They divided 256 intensity levels into 16 scales, and defined the brightness at the peaks of each color channel to 12/16, 6/16, and 2/16 for red (R), green (G), and blue (B), respectively. Results in the color balance yielded at the band ratios of $G/R = 0.5$ and $B/R = 0.17$.

In this paper, we propose an improvement of the previously discussed methods to increase the visibility of a hidden anatomical structure in dark regions for non-uniform illumination retinal images. Gamma correction is firstly used to improve image luminosity. The image contrast is then enhanced by applying multi-scale tone manipulation, and finally, color balance is adjusted by specifying the image brightness based on Hubbard's model.

The performance of the proposed method was evaluated by using the publicly available datasets, the Diabetic Retinopathy Database (DIARETDB1) [14]. The

DIARETDB1 dataset, acquired by Kauppi et al. [14], consists of 89 color retinal images. The images were taken with a 50° field of view with the size of 1500×1152 pixels and stored in a 24-bit PNG format.

The results demonstrate that the proposed algorithm performs well in non-uniform illumination image enhancement. Images processed by our method have better visual quality and are better suited to human visual perception than those processed by NCNR and LCA. This will allow ophthalmologists to more visually and accurately diagnose the disease.

The paper proceeds as follows: the proposed method is described in Section 2, our experimental results appear in Section 3, and conclusions in Section 4.

2 CONTRAST AND COLOR BALANCE ENHANCEMENT

The proposed method consists of three modules (see Fig. 1): the Luminosity enhancement module, the Contrast and tone enhancement module, and the Color balance adjustment module. The luminosity enhancement module employs gamma correction for boosting up the visibility of anatomical details that were obscure in dark regions. In the second module, the Multi-scale tone manipulation is employed to improve the image contrast and tone. The color balance adjustment in the last module is performed to specify the brightness of red, green, and blue channel based on the Hubbard specification. A detailed description of each module is presented in the following subsections.

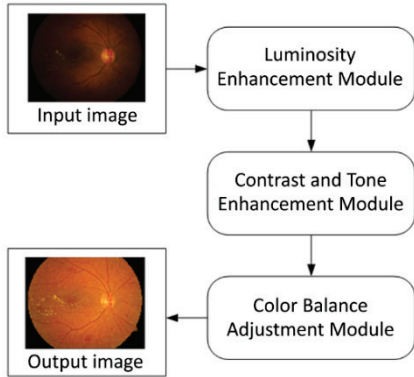


Figure 1 Framework of the proposed method.

2.1 Luminosity Enhancement Module

From our observation in the DIARETDB1 datasets, most images have some dark regions caused by uneven luminance. Those regions obscure the anatomical structures in the ocular fundus, and make the details necessary to make a diagnosis undetectable. To improve those dark regions of the image, the luminosity of the image needs to be enhanced. Mei et al. introduced a luminance gain matrix [7], which was derived from the gamma correction of the value channel in the HSV color space. The gain matrix showed good improvement in luminosity. The gain $G(x,y)$ could be rewritten in the simple form as:

$$G(x,y) = \frac{m \left[\frac{v(x,y)}{m} \right]^{\frac{1}{2.2}}}{v(x,y)}, \quad (1)$$

where v refers to the luminance values from HSV, that is, $v(x,y) = \max(r(x,y), g(x,y), b(x,y))$ and $r(x,y)$, $g(x,y)$, and $b(x,y)$ are intensity values at the coordinate (x,y) from the R, G, and B channels, respectively. The parameter m is the maximum value of v .

In particular, from our investigation in the DIARETDB1 dataset, the average dynamic range of red, green, and blue channels are [0 – 222], [0 – 157], and [0 – 78], respectively. Due to the dynamic range of the blue channel being narrow and its maximum intensity value being low, hence, the formula of $v(x,y) = \max(r(x,y), g(x,y), b(x,y))$ could be changed to $v(x,y) = \max(r(x,y), g(x,y))$ to reduce the computational time. Additionally, the parameter m might be estimated from the maximum intensity of the red channel as $m = \max(r(x,y))$; in this case, the parameters v and m can be a parallel approximation.

The enhanced luminosity RGB' , from the input image, RGB , can be obtained by multiplying pixel by pixel as in the following:

$$RGB'(x,y) = RGB(x,y) \cdot G(x,y). \quad (2)$$

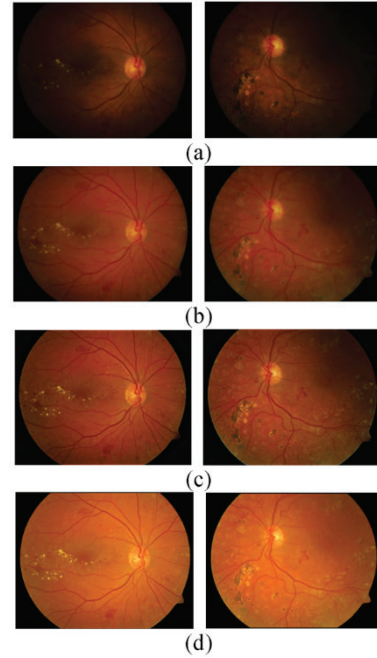


Figure 2 Results from each enhancement module: (a) Input images from the file Image007.png (left column) and Image027.png (right column); (b) Luminosity Enhancement; (c) Contrast and tone Enhancement; (d) Color balance adjustment.

The second row of Fig. 2 shows the enhanced luminosity as seen in the images; visibility is improved in the dark areas. The next step is to enhance the image contrast and tone to provide a better visibility of the retinal anatomical structures.

2.2 Contrast and Tone Enhancement Module

Multi-scale tone manipulation is effectively used for improving the contrast and tone of images. In this module, the image is converted to the $L^*a^*b^*$ color space, while lightness, the L component, is decomposed with different scales [15]. In our scheme, the decomposed image consists of a smooth base layer (l_1) and two detail layers (d_0 and d_1). Each layer is processed separately and recombined to generate the final result [15].

Farbman et al. [15] introduced a smooth layer, l_1 , which can be obtained by using the WLS filter. When varying the parameter values of the WLS filter, the image tone is not smooth enough, especially, around the optic disk areas, and it results in an unnatural look as shown in Fig. 3(b). Hence, the WLS is replaced with the optic transfer function of Deeley et al. [16], and named as DOTF. The smoother lightness versions l_0 and l_1 can be defined as:

$$l_0 = F^{-1}\{F\{l\} \cdot DOTF\}, \quad (3)$$

$$l_1 = F^{-1}\{F\{l_0\} \cdot DOTF\}, \quad (4)$$

where F and F^{-1} denote the Fourier transform and its inverse, and the formula of DOTF described in [16] is given as:

$$DOTF = \exp\left[-\left(\frac{u}{(20.9 - 2.1\Phi)^{(1.3-0.07\Phi)}}\right)\right], \quad (5)$$

where u denotes the spatial frequency and Φ is the pupil diameter approximated to 4 mm.

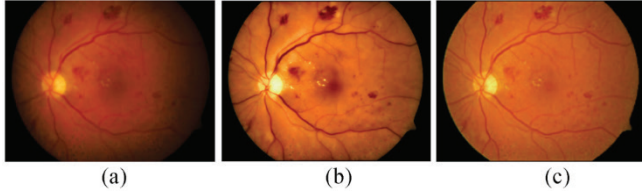


Figure 3 Comparison of results with a different filter: (a) Input image from the file Image008.png; (b) Output by using WLS; (c) Output by using OTF.

The lightness channel l and two smoother lightness versions l_0 and l_1 are then used to generate two detail layers, d_0 and d_1 , expressed in the form:

$$d_0 = S(\delta_0, l - l_0), \quad (6)$$

$$d_1 = S(\delta_1, l_0 - l_1), \quad (7)$$

where S is a sigmoid curve which is defined as $S(a, x) = 1/(1+\exp(-ax))$. δ_0 and δ_1 are boosting factors for the d_0 and d_1 layers.

The base layer l_1 and two detail layers, d_0 and d_1 were then recombined to form the enhanced lightness, l' , which can be formulated as:

$$l' = \mu + S(\delta_2, \eta \cdot l_1 - \mu) + d_0 + d_1, \quad (8)$$

where η denotes the exposure of the base layer, δ_2 refers to the boosting factors for the base layer, and μ is the mean of the lightness range. We utilized parameters in agreement with those reported by Farbman et al. [15]. The parameters were $\delta_0 = 1$, $\delta_1 = 40$, and $\delta_2 = 1$, $\eta = 1.0$ and $\mu = 56$.

In the last step of this module, the contrast and tone of the enhanced image, RGB'' , are obtained by replacing the L component by l' and converting $L^*a^*b^*$ back to the RGB color space.

Fig. 4 shows the base layer and two detail layers corresponding to the image in Fig. 2. DOTF takes less processing time than WLS and gives a more natural look as shown in Fig. 3(c). The contrast and tone of the enhanced images in the third row of Fig. 2 have a better contrast when compared to the luminosity enhanced images in the second row. The next module is to adjust the image color balance.

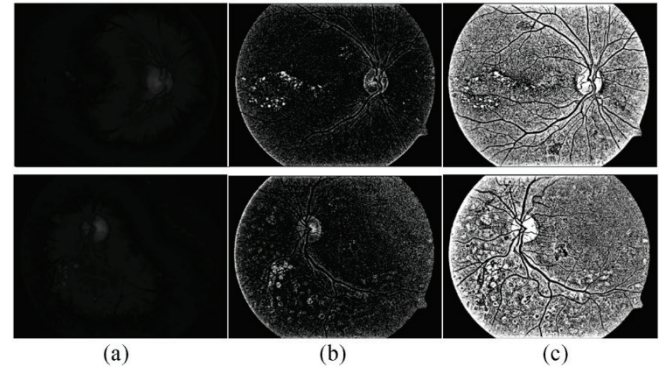


Figure 4 Corresponding base layer and two detail layers of the image in Fig. 2: (a) Base layer l_1 , (b) and (c) Detail layers d_0 and d_1 .

2.3 Color Balance Adjustment Module

The findings by Hubbard et al. [13] lead to the main motivation for adjusting the color balance of color retinal images. The experimental results in the AREDS2 reading center confirm that a colored image with Hubbard's model parameters shows a greater contrast of drusen and pigment abnormalities compared to the normal retinal pigment epithelium. They specified the color balance with the color brightness of R , G , and B to 192, 96, and 32, respectively, which resulted in color ratios of $G/R = 0.5$ and $B/R = 0.17$.

In this module, we use the specified brightness vector, $sb_b = [192, 96, 32]$ from Hubbard's model to adjust the color balance of the enhanced result, RGB'' , as processed in the above mentioned module. The color offset of the image RGB'' is adjusted by the difference between the brightness vector and the mean vector, μ , of the RGB'' image, that is $\mu = \text{mean}(RGB'')$.

The color balance for the final output image of RGB''' is formulated as:

$$RGB'''(x, y) = RGB''(x, y) + (sb_b - \mu), \quad (9)$$

The fourth row of Fig. 2 shows the results of the color balance adjusting module which give a greater contrast of the foreground compared to the background.

3 EXPERIMENTAL RESULTS

In this section, the output results from our method are compared with NCNR [9] and LCA [7] by using the images from the DIARETDB1 dataset.

To evaluate the performance of our method, various quantitative metrics including measuring colorfulness ($M^{(3)}$) [17], visual saliency-based index (VSI) [18], and lightness order error (LOE) [19] are used to measure image quality. $M^{(3)}$ [17] is a colorfulness metric which is fitted to the perceptual data collected from a psychophysical experiment and it has a simple expression based on the opponent color. VSI [18] measures the image quality in consistent with the human visual system or subjective evaluations. The method uses the visual saliency map as a feature to measure the image's local quality. LOE [19] measures the lightness order error between the original and output image to assess naturalness preservation. The smaller the LOE values, the better that naturalness is preserved. However, color image

assessment is difficult to evaluate by using only quantitative measurements. To determine if an image has good quality requires both quantitative and visual assessment. For the visual assessment, we visually inspected the output image to see if it retains a pleasing natural look and if lesions of the ocular diseases can be easily discriminated.

3.1 Visual Assessment

Visual assessments are shown in Fig. 5 with the images taking from DIARETDB1 where the first column represents the original images, and the output images consist of the proposed method in the second column, NCNR in the third column, and LCA in the fourth column. The output images, by employing the NCNR and LCA methods, generated dark color areas, especially for the blood vessel, macular, and the abnormal tissue such as hemorrhages, and small red blood dots.

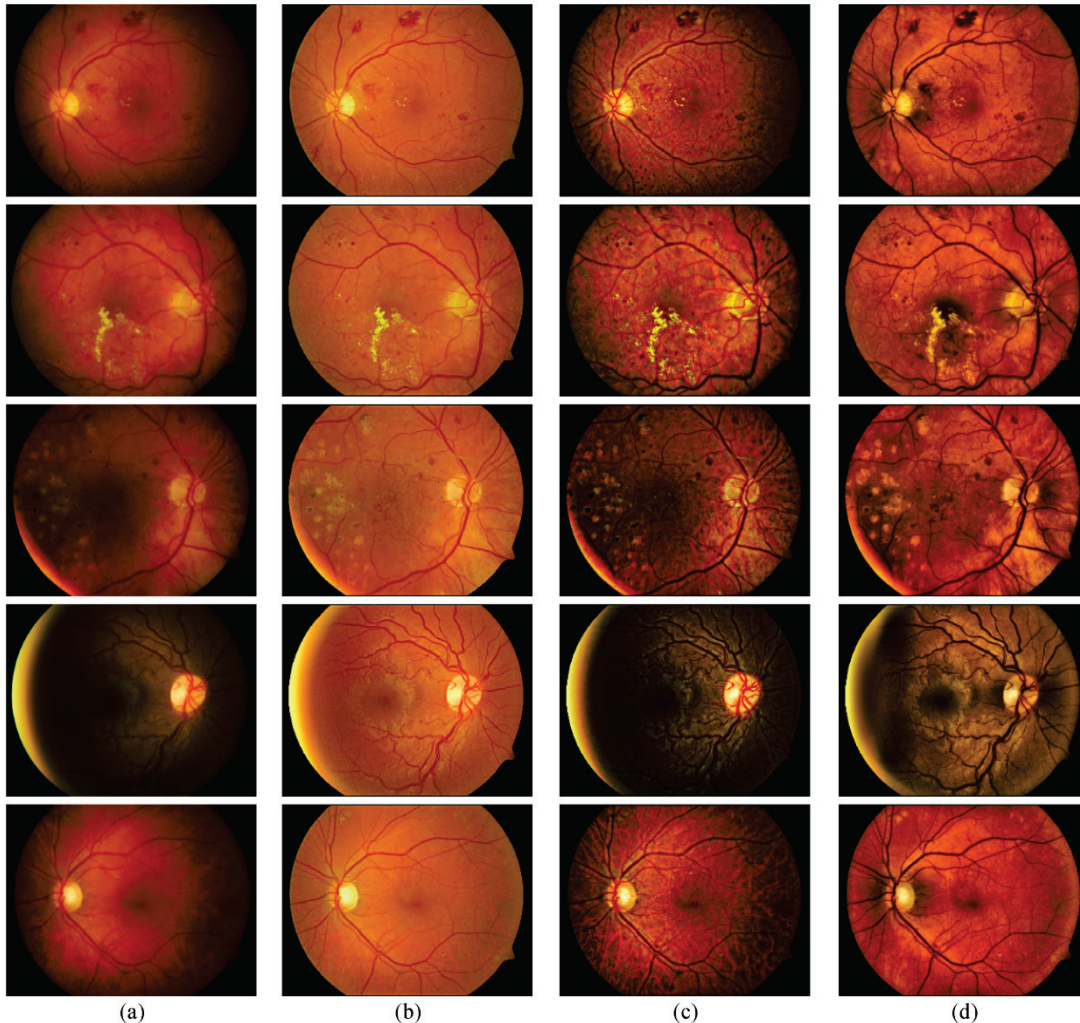


Figure 5 Results of different methods for the images: image008.png, image019.png, image029.png, image047.png, and image054.png from DIARETDB1: (a) Input images and output images, (b) the proposed method, (c) NCNR and (d) LCA.

Although the proposed method cannot perform image contrast as well as VCEA and LCA, our proposed method produces a pleasing visual appearance and maintains the

color balance better than the other methods. The output results from the proposed method could reveal the hidden

anatomical structures from the dark regions and may help ophthalmologists to improve their diagnostic accuracy.

Compared with the other methods, the proposed method enhances the contrast and tone of anatomical details. Additionally, it also produces a natural look with a proper color balance as specified by the reading center [13].

3.2 Quantitative Assessment

Tab. 1 shows all quantitative metric values, where each row corresponds to the enhancement methods and each column corresponds to the quantitative metrics in the form of an average and standard deviation value (Mean \pm SD), calculated from 89 images from DIARETDB1. Illustrated are

Table 1 Performance of various methods on DIARETDB1

Method	G/R	B/R	$M^{(3)}$	VSI	LOE
Our method	0.50 ± 0.00	0.17 ± 0.00	97.11 ± 1.73	0.9574 ± 0.0149	97.39 ± 35.15
NCNR	0.39 ± 0.06	0.09 ± 0.05	72.23 ± 19.27	0.9539 ± 0.0100	134.77 ± 29.17
LCA	0.39 ± 0.06	0.08 ± 0.05	86.47 ± 18.57	0.9556 ± 0.0121	474.66 ± 116.02

4 CONCLUSION

In this paper, a new automatic image enhancement method is proposed to improve contrast and color balance for the non-uniform illumination retinal image. Our proposed method was tested on the DIARETDB1 dataset. The quantitative and visual assessment results demonstrate that the output images obtained by the proposed method give natural look and therefore are better suited for human visual perception in comparison to other methods. In terms of diagnoses made by ophthalmologists, the enhanced images obtained by our method could be used to assist ophthalmologists in the early detection of a disease and its diagnosis.

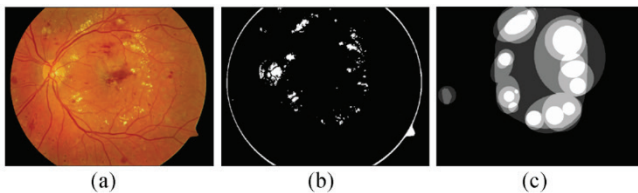


Figure 6 Example of a hard exudate detection: (a) Enhanced image of the image005.png; (b) Segmented result; (c) Hard exudate ground truth.

5 FUTURE WORK

Due to the output images derived from our proposed method they are standardized in color as shown in Fig. 5, hence, color features such as the color ratios: G/R and B/R could be used to automatically detect the presence of abnormal lesions such as hard exudate, hemorrhage, and small red blood dot. Fig. 6 shows an example of hard exudate detection by thresholding only the G/R color ratio feature. Fig. 6(b) depicts the hard exudate segment as a result from the enhanced input image in Fig. 6(a). To increase the accuracy of the detection, the optic disk area should be removed from the retinal color image before detection and an effective SVM classification should be applied to classify the

the mean and standard deviation of the color ratios, $M^{(3)}$, VSI , and LOE . From the experimental results in the column two and three of Tab. 1, the average green-to-red, and blue-to-red ratio of the output image from the proposed method were 0.50 ± 0.00 , 0.17 ± 0.00 , which meets the color balance specification from the Hubbard model. The proposed method gave the highest quantitative scores for $M^{(3)}$ (97.11 ± 1.73) and VSI (0.9574 ± 0.0149). These highest scores imply that the proposed method yields better colorfulness and visual perception. The experimental results in the last column of Tab. 1 show that the proposed method scored the lowest for LOE (97.39 ± 35.15), and therefore can well preserve the naturalness.

hard exudate instead of using the threshold; the classification results are shown in Fig. 6(c).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Prince of Songkla University Grant no. SCI600372S.

Remark

The article was orally presented at the 23rd International Computer Science and Engineering Conference (ICSEC2019).

6 REFERENCES

- [1] Main causes of visual impairment. *WHO*. Retrieved from <http://who.int/blindness/causes/priority/en>
- [2] Wong, W. L., Su, X., Li, X., & Cheung, C. M. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *The Lancet Global Health*, 2(2), 106-116. [https://doi.org/10.1016/S2214-109X\(13\)70145-1](https://doi.org/10.1016/S2214-109X(13)70145-1).
- [3] Abramoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE Rev in Biomed Eng*, 3, 169-208. <https://doi.org/10.1109/RBME.2010.2084567>
- [4] Wang, C., & Ye, Z. (2005). Brightness preserving histogram equalization with maximum entropy: a variational perspective. *IEEE Trans. Consum. Electron.*, 51(4), 1326-1334. <https://doi.org/10.1109/TCE.2005.1561863>
- [5] Ibrahim, H. & Kong, N. (2007). Brightness preserving dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.*, 53, 1752-1758. <https://doi.org/10.1109/TCE.2007.4429280>

- [6] Sheet, D., Garud, H., Suveer, A., Mahadevappa, M., & Chatterjee, J. (2010). Brightness preserving dynamic fuzzy histogram equalization. *IEEE Trans. Consum. Electron*, 56(4), 2475-2480. <https://doi.org/10.1109/TCE.2010.5681130>
- [7] Mei, Z., Kai, J., Shaoze, W., Juan, Y., & Dahong, Q. (2017). Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Rev. Biomed. Eng.* (99), 1-7. <https://doi.org/10.1109/TBME.2017.2700627>
- [8] Zuiderveld, K. (1994). *Contrast limited adaptive histogram equalization*. San Diego: Academic press professional, 474-485. <https://doi.org/10.1109/ICACCI.2014.6968381>
- [9] Peishan, D., Hanwei, S., Jianmei, Z., Ling, L., Jing, W., & Min, F. (2016). Retinal fundus image enhancement using the normalized convolution and noise removing. *International Journal of Biomedical Imaging*, 1-13. <https://doi.org/10.1155/2016/5075612>
- [10] Gastal, E. S., & Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. *ACM Trans. on Graphics*, 30(4), 1244-1259. <https://doi.org/10.1145/2010324.1964964>
- [11] Rajan, J., Kannan, K., & Kaimal, M. R. (2008). An improved hybrid model for molecular image denoising. *J. of Math. Imaging and Vision*, 31(1), 73-79. <https://doi.org/10.1007/s10851-008-0067-4>
- [12] Hamza, A. B., Luque-Escamilla, P. L., Martinez-Aroza, J., & Roman-Roldan, R. (1999). Removing noise and preserving details with relaxed median filters. *J. of Math. Imaging and Vision*, 11(2), 161-177. <https://doi.org/10.1023/A:100839551>
- [13] Hubbard, L. D., Danis, R. P., Neider, M. W., Thayer, H. D., Wabers, H. D., White, J. K., . . . Pugliese, M. F. (2008). Brightness, contrast, and color balance of digital versus film retinal images in the age-related eye disease study 2. *Journal of Investigative Ophthalmology & Visual Science*, 49(8), 3269-3282. <https://doi.org/10.1167/iovs.07-1267>
- [14] Tomi, K., Valentina, K., Kristian, K. J., Lasse, L., & Liris, S. (2007). Diaretdb1 diabetic retinopathy database and evaluation protocol. in *Proc. of the British Machine Vision Conference*. University of Warwick, UK, 252-261. <https://doi.org/10.5244/C.21.15>
- [15] Farbman, z., Fattal, R., Lischinski, D., & Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM TOG*, 27(3), 67. <https://doi.org/10.1145/1360612.1360666>
- [16] Deeley, R. J., Drasdo, N., & Charman, N. W. (1991). A simple parametric model of the human ocular modulation transfer function. *Phthalmic & Physiological Optics*, 11(1), 91-93. <https://doi.org/10.1111/j.1475-1313.1991.tb00200.x>
- [17] Hasler, D. & Susstrunk, S. (2003). Measuring colorfulness in natural images. *Proc. SPIE*, 5007, 87-95. <https://doi.org/10.1117/12.477378>
- [18] Zhang, L., Shen, Y., & Li, H. (2014). VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans Image Process*, 23(10), 4270-4281. <https://doi.org/10.1109/TIP.2014.2346028>
- [19] Wang, S., Zheng, J., Hu, H., & Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *Trans. Image Process.*, 22(9), 60-70. <https://doi.org/10.1109/TIP.2013.2261309>

Authors' contacts:

Preecha VONGHIRANDECHA

Prince of Songkla University,
15 Kanjanavanich, 90110, Hatyai, Songkhla, Thailand
preecha.v@psu.ac.th

Montri KARNJANADECHA, PhD

Prince of Songkla University,
15 Kanjanavanich, 90110, Hatyai, Songkhla, Thailand
montri@coe.psu.ac.th

Sathit INTAG, PhD

Prince of Songkla University,
15 Kanjanavanich, 90110, Hatyai, Songkhla, Thailand
sathit.i@psu.ac.th

HETEROGENEOUS FENTON-LIKE DECOLORIZATION OF PROCION RED MX-5B WITH IRON-ALGINATE GEL BEADS AS AN EFFECTIVE CATALYST

Memduha ERGÜT, Ayla ÖZER

Abstract: In the present study, iron-alginate gel beads (Fe-Alg gel beads) were synthesized by the entrapment method and subsequently, the prepared gel beads were characterized by FTIR, SEM, and EDX analysis methods. The characterization studies showed that Fe-Alg gel beads were of spherical morphology and iron was successfully entrapped in alginate. Subsequently, the synthesized Fe-Alg gel beads were used as a heterogeneous catalyst for the decolorization of a hazardous azo-dyestuff, Procion Red MX-5B (PR MX-5B), with Fenton-like decolorization. The effects of process parameters such as initial pH, H₂O₂ concentration, initial dye concentration and catalyst concentration on Fenton-like decolorization were investigated. For Fenton-like decolorization of PR MX-5B, the optimum process parameters were determined to be pH: 3.0, 20 mμ of H₂O₂ concentration, 50 mg/L of initial dye concentration, and 5.0 g/L of catalyst concentration, respectively. The reaction kinetics was well fitted to second order reaction kinetics. As a result, the synthesized gel beads have a high color removal efficiency for PR MX-5B dye from an aqueous solution.

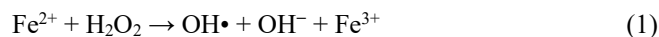
Keywords: decolorization; gel beads; heterogeneous Fenton-like catalyst; iron-alginate; Procion Red MX-5B

1 INTRODUCTION

Nowadays, dyes are utilized in approximately all industrial sectors; therefore, dye-laden wastewaters contaminate natural waters and reduce water's appreciate in use [1]. For this reason; several physicals, chemical and biological treatment methods such as adsorption, ion-exchange, photocatalysis, membrane separation, coagulation, electrocoagulation, and wet air oxidation have been tested to remove dye effluents and many of these techniques have their advantages and disadvantages. These advanced processes are not always efficient enough since the removal of dyes are difficult due to their water solubility and high molecular strength. Moreover, they have suffered from the generation of toxic chemical sludges as a result of the transfer of pollution from the wastewater to the solid phase, high operational costs of some and their limited applicability [2, 3].

Recently, advanced oxidation processes (AOPs) such as Fenton, photo-Fenton processes, electro-Fenton, ozonation, electrochemical oxidation, photolysis with H₂O₂ and O₃, and UV/H₂O₂ processes have attracted considerable attention as they are alternative methods without sludge formation for the treatment of a wide variety of organic contaminants in wastewaters. Their degradation mechanism mainly depends on the generation and the oxidative behavior of hydroxyl radicals [4]. The complicated organic molecules could be either oxidized to smaller organics or entirely mineralized to carbon dioxide (CO₂) and water (H₂O) by OH• radicals under treatment by AOPs [5].

The Fenton-like process which is the most hopeful method among the AOPs is a strong source of hydroxyl radicals from H₂O₂ in the presence of iron cations at an acidic medium. Conventional homogeneous Fenton's oxidation includes the implementation of ferrous or ferric salts and hydrogen peroxide to generate hydroxyl radicals. The main principle of oxidation was presented in Eqs. (1)-(2):



The traditional homogeneous Fenton process is a simple process owing to operating at room temperature and atmospheric pressure. Additionally, it is quite effective when the wastewater contains suspended solids at high levels due to limited photon penetration. However, the important drawbacks of this process are narrow working pH range that is reported and that generally results in higher degradation yield at around pH 3.0, the trouble of separation and recovery of the homogeneous catalyst and the formation of the large volume of iron sludge.

Therefore, to get rid of this problem, the usage of heterogeneous Fenton catalysts has been improved, such as iron-containing metal oxides, iron-containing bimetallic nanoparticles, iron-bearing porous materials such as clays, resins, zeolites, mesoporous substances, and iron species supported by neutral organic polymers (i.e., sodium alginate, sodium carboxymethyl cellulose) and inorganic materials [6 – 8].

Sodium alginate (SA) is one of the extensively studied natural polysaccharides to produce stable gels owing to its reaction ability with many cations such as Fe³⁺, Ca²⁺, and Ba²⁺ [9]. In the present study, iron was entrapped in sodium alginate, which is a biopolymer matrix, to prevent producing sludge at the end of the reaction. The entrapment method within alginate beads is one of the most extensive, effective, simple and inexpensive techniques for the entrapment of surfactants, activated carbon, and especially metal hydroxides (i.e., Fe³⁺ and Ni²⁺) for wastewater treatment. The benefits of entrapment within alginate beads are non-toxicity and biodegradability as well as their porosity facilitates interaction with the entrapped material diffusion of solutes into beads. Moreover, these gels can be separated from the

reaction medium by a simple filtration owing to their water-insoluble properties [10].

The objective of the presented study was synthesis and characterization of Fe-alginate gel beads and investigation of the applicability of them as a heterogeneous catalyst for the Fenton-like decolorization of a textile azo-dye, Procion Red MX-5B, in a batch system. For this purpose, the main process parameters such as the concentration of H_2O_2 , initial pH of the solution, initial dye concentration and catalyst (Fe-Alg gel beads) concentration on Fenton-like decolorization were investigated.

2 MATERIAL AND METHODS

2.1 Materials and Preparing of the Solutions

The monoazo dye Procion Red MX-5B (PR MX-5B), also called the Reactive Red 2, was supplied by Sigma-Aldrich. PR MX-5B was of commercial purity (Type: Anionic, $M_w = 615.33$ g/mol, $\lambda_{\max} = 538$ nm). PR MX-5B has an azo double bond as a member of the chromophoric group. The chromophoric groups absorb light and affect visibility in water bodies [11]. PR MX-5B has an elemental composition of $\text{C}_{19}\text{H}_{10}\text{Cl}_2\text{N}_6\text{Na}_2\text{O}_7\text{S}_2$, and its chemical structure is presented in Fig. 1.

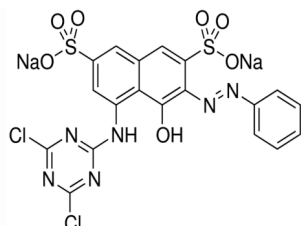


Figure 1 Chemical structure of PR MX-5

FeCl_3 , $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$, and H_2O_2 (30%) were provided from Merck. Sodium alginate ($\text{C}_6\text{H}_9\text{NaO}_7$) was purchased from Sigma Aldrich.

The stock PR MX-5B solution of 1000 mg/L was prepared by dissolving 1 g of dye in 1 L of deionized water; then the desired PR MX-5B concentrations were prepared by proper dilutions from stock dye solution. 10 mM, 20 mM and 30 mM of H_2O_2 solutions were prepared by dilution from the 50 mM of stock H_2O_2 solution which was prepared from the 30.0% H_2O_2 solution (9.73M). The solution of sodium alginate 2.0% (w/v) was prepared by dissolving the required weight of sodium alginate in 100 mL deionized water. 0.15 M Ba^{2+} and 0.05 M Fe^{3+} solutions were prepared by dissolving an accurate weight of $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$ and FeCl_3 in 100 mL of distilled water, respectively.

2.2 The Synthesis of Iron-Alginate Gel Beads

Iron-alginate gel beads were synthesized by the method previously reported by Rosales et al., 2012 [10]. According to that, 100 mL of sodium alginate solution 2.0% (w/v) was dropped into 200 mL of the solution composed of 0.15 M Ba^{2+} and 0.05 M Fe^{3+} in 1:1 volume ratio by a syringe pump at a flow rate of 5 mL/min into for the gel bead formation.

The spherical iron-alginate gel beads (Fe-Alg gel beads) formed by the contact of the alginate including Fe^{3+} and Ba^{2+} ; after that, the solution containing beads was stirred magnetically at a low velocity and 25 °C temperature for 2 h for the stabilization of beads. The formed brown particles (Fig. 2) were treated at +4 °C for 2 h in the gelling solution. Thereafter, they were simply decanted and washed with distilled water. Finally, the beads were dried in an oven at 70 °C for 4 h.

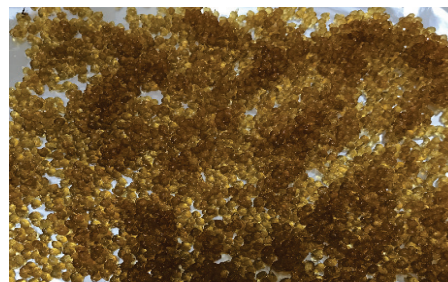


Figure 2 The prepared iron-alginate gel beads

2.3 Characterization Studies

The morphology and elemental identification of beads were analyzed by the Scanning Electron Microscope (SEM) analysis, and Energy Dispersive X-ray Analysis (EDX), respectively before and after decolorization experiments. The functional groups of synthesized Fe-Alg gel beads were determined by the Fourier Transform Infrared Spectrometer (FT-IR) in the range of $4000 - 400$ cm^{-1} , pre- and post- decolorization.

2.4 Heterogeneous Fenton-Like Decolorization Experiments

Heterogeneous Fenton-like decolorization experiments of PR MX-5B were carried out on 100 mL of dye solutions in a water bath at a constant shaking rate and constant temperature (25°C). In decolorization experiments, the required amount of Fe-Alg gel beads was added to the dye solution which was adjusted to the desired initial concentration and initial pH value. Then, 5.0 mL of H_2O_2 solution at different concentrations were added rapidly to the PR MX-5B dye solution and the Fenton-like decolorization began. The Erlenmeyer flasks (250 mL) were agitated water baths for the required reaction times. Then, samples were taken at pre-determined time intervals. Afterward, the UV-vis absorbance readings were carried out on samples by a spectrophotometer (Specord 210 Plus, Analytic Jena, Germany).

The decolorization percentage for PR MX-5B was reduced in UV-vis absorbance and determined by Eq. (3) as follows:

$$D = \frac{A_0 |_{538 \text{ nm}} - A_t |_{538 \text{ nm}}}{A_0 |_{538 \text{ nm}}} \times 100 \% \quad (3)$$

Where D (%) represents decolorization efficiency, A_0 and A_t are the absorbance values at an initial time and at any time of

PR MX-5B. Moreover, the control tests without an added catalyst were done to detect the decolorization capacity of H_2O_2 for PR MX-5B.

Experiments were carried out to examine the effects of other Fenton-like reaction parameters on dye decolorization as pH (3 – 6), initial dye concentration (100 – 500 mg/L), Fe-Alg gel beads (catalyst) concentration (0.5 – 5 g/L) and H_2O_2 concentration.

3 RESULTS AND DISCUSSION

3.1 The Characterization of Iron-Alginate Gel Beads

The morphology and elemental analysis of the synthesized gel beads were investigated by SEM and EDX analysis, respectively, before and after decolorization. The SEM images are shown in Fig. 3 (a, b, c, d). As seen in Fig. 3a, the beads have a spherical shape and the SEM image (Fig. 3d), after the Fenton-like reaction showed that the surface change may be due to breakage of alginate chains and iron ions came out from the inner surface of beads. Furthermore, the average particle size of dried iron-alginate gel beads was calculated to be 231 μm by using the Image J program with SEM images.

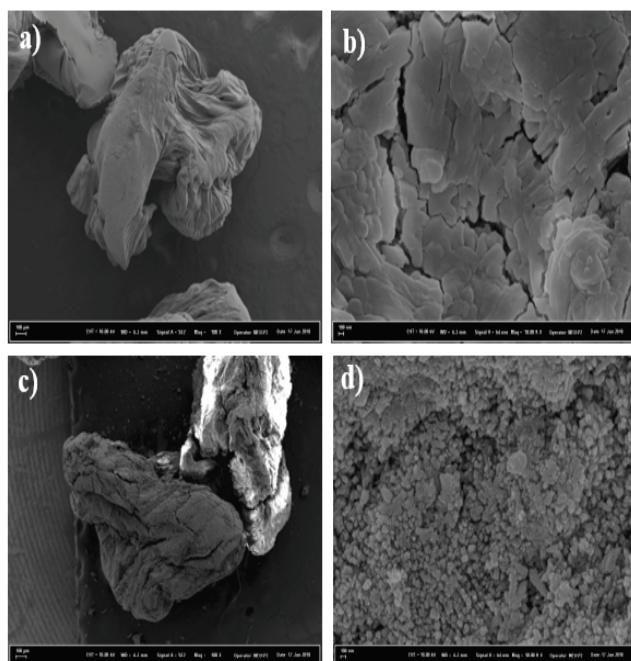


Figure 3 SEM images of iron-alginate gel beads before reaction: a) 100 KX, b) 50.00 KX; iron-alginate gel beads after reaction: c) 100 KX, d) 50.00 KX

The EDX spectrums and the elemental and quantitative weight composition (wt. %) of Fe-Alg gel beads before and after the reaction were presented in Fig. 4 (a, b) and Tab.1, respectively. The cross-link between Fe and alginate leads to obtaining brown beads according to alginate beads without Fe. It is reported that ionic cross-links occur when alginate gets in touch with Ca^{2+} and Ba^{2+} to create an "egg-box" structure.

Hence, when alginate droplets are immersed into a ferrous aqueous solution, ionic cross-links could be

constructed among the carboxyl group on alginate chains and iron [10].

According to the EDX results, the prepared Fe-Alg gel beads involve the elements of Fe, Ba, C, O, Cl, Pt, Pd, Na, N, S, and Al (the peaks of the elements of Pt and Pd stem from the coating of the sample) before reaction. As a consequence, the formation of Fe-Alg gel beads was approved by the signals in the Fe and Ba regions. Furthermore, the elements of C, O, and Na stem from the alginate structure ($\text{C}_6\text{H}_7\text{NaO}_6$)_n. From the EDX spectrum, after the reaction, Fe, Ba, C, O, Cl, Pt, Pd, Na, N, and S elements were detected in the structure.

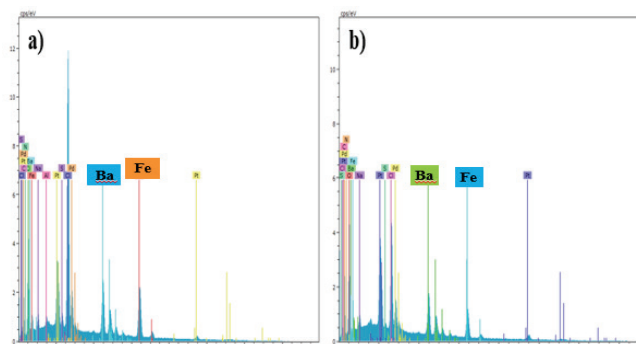


Figure 4 EDX spectrums of Fe-Alg gel beads a) before b) after the reaction

Table 1 Elemental composition of Fe-Alg gel beads before and after the Fenton-like degradation

Element	Weight (%)	
	Synthesized Fe-Alg gel beads	After the Fenton-like reaction
Ba	24.65	26.07
Fe	24.44	17.99
C	6.45	16.80
O	17.79	24.90
Cl	22.50	11.97
Na	1.678	1.523
N	1.115	1.421
S	0.716	0.076
Al	0.657	—

The FT-IR spectra of Fe-Alg gel beads before and after decolorization were presented in Fig. 5 (a, b). In Fig. 5.a, the broad peak at around 3300 cm^{-1} is ascribed to the hydrogen-bonded O–H stretching, and 2362 cm^{-1} is due to the $\text{C}\equiv\text{C}$ stretching vibration [10, 12]. The band at 1028 cm^{-1} is attributed to the C–O–C stretching due to the saccharide structure of sodium alginate [13]. The band at 1406 cm^{-1} is contributed to the C–OH deformation vibration with the contribution of the O–C–O symmetric stretching vibration of the carboxylate group. 1596.6 cm^{-1} is assigned to the asymmetric stretching vibrations of the carboxyl group of the alginate molecule. The peak at 1726 cm^{-1} in the spectrum of the Fe-Alg gel bead also belongs to the absorption vibrations of the carboxyl group of the alginate molecule [10, 13].

Moreover, as seen from Fig. 5, a decrease was observed in the FT-IR peaks' intensity in the part before and after degradation. This situation may result from the change in the distance of the C–O bond of the carboxylate group due to the bonding strength between the iron ion and the oxygen of the

carboxyl group changes [10]. As seen from Fig. 5, b, the peak at 1726 cm^{-1} disappeared after decolorization. This result could be due to a decrease in the absorption vibrations of the carboxyl group of the alginate molecule.

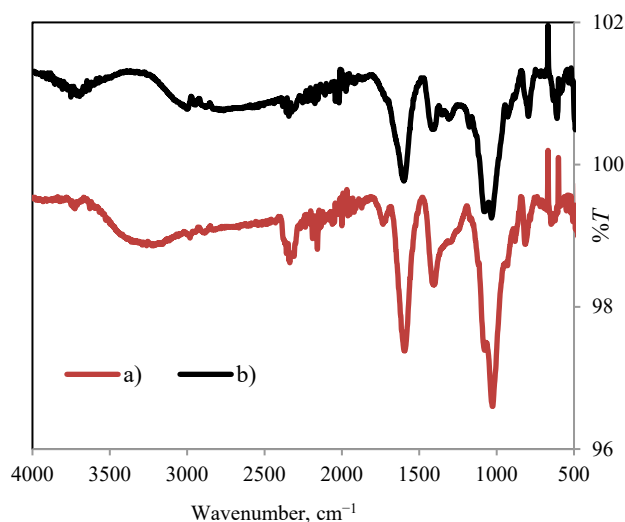
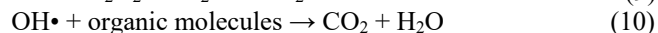
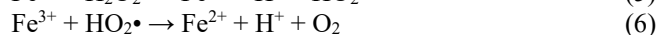
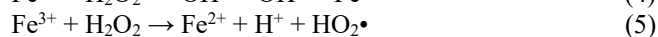
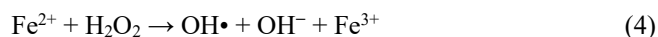


Figure 5 FT-IR spectra of Fe-Alg gel beads a) before b) after the Fenton-like reaction

3.2 The Analysis of Dye Decolorization

The catalytic efficiency of the synthesized Fe-Alg gel beads was assessed in the heterogeneous Fenton-like decolorization of PR MX-5B.

The Fenton degradation mechanism depends on the production of reactive hydroxyl radicals ($\text{OH}\cdot$) by the decomposition of H_2O_2 in the presence of Fe^{2+} ions. In the classical Fenton system, the formation cycles of hydroxyl radicals were given in Eqs. (4) – (10); [14].



According to Eqs. (4) – (10), firstly, ferrous ions begin the Fenton-like oxidation and cause the formation of hydroxyl radicals, and afterward, these form hydroxyl radicals and assault the organic contaminants and lead to their disruption to smaller organic compounds [14].

The progress of the Fenton-like decolorization of the PR MX-5B dye was observed by the UV-vis spectral analysis. The variations in the spectrum over time were presented in Fig. 6.

Moreover, the changes in the color of dyestuff solution overtime were given in the inset of Fig. 6. PR MX-5B has the absorption peaks at 510.97 nm and 538.04 nm, which are induced by the $\pi \rightarrow \pi^*$ electronic transition in a conjugated system linked by azo groups [15].

As seen in Fig. 6, adsorption peaks progressively declined and almost disappeared with the increase of reaction time. Furthermore, no new peaks that formed during the reaction were determined.

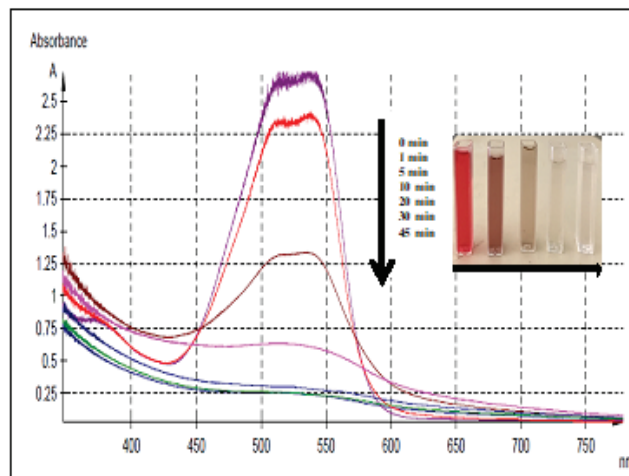


Figure 6 The UV-vis spectra of the PR MX-5B dye after decolorization ($C_0 = 100\text{ mg/L}$, $C_{\text{H}_2\text{O}_2} = 20\text{ mM}$, $X_0 = 1\text{ g/L}$, $\text{pH} = 3.0$, $T = 298\text{ K}$)

Consequently, the chromophoric group ($\text{N}=\text{N}$ group), as well as the conjugated system in the PR MX-5B molecular structure, was broken and finally, the PR MX-5B molecule was decolorized.

3.3 Effects of Parameters on the Decolorization of Procion Red MX-5B

3.3.1 Effect of the hydrogen peroxide concentration

H_2O_2 concentration is a fundamental parameter to be determined in the heterogeneous Fenton-like reactions because of number of $\text{OH}\cdot$ radicals which will initiate the reaction, depends directly on H_2O_2 concentration. At low concentrations of H_2O_2 , not enough hydroxyl radicals can be obtained for efficient degradation, which is why the degradation rate could be long-term. Moreover, the best part of free radicals is immediately used up by the dye. Conversely, more $\text{OH}\cdot$ radicals could be produced in the presence of a high concentration of H_2O_2 , but these formed $\text{OH}\cdot$ radicals preferably react with the excess of H_2O_2 [1].

The effect of the H_2O_2 concentration was presented in Fig. 7, at the experimental conditions of initial dye concentration: 100 mg/L ; natural pH of the solution (nearly 5.2), catalyst concentration 1 g/L , and temperature 298 K .

As seen from Fig. 7, quite close ($\approx 98\%$) decolorization percentages were acquired when the H_2O_2 concentrations were 20 mM , 30 mM , and 50 mM . Hence, the optimum H_2O_2 concentration was selected as 20 mM in order to prevent unprofitable consumption of H_2O_2 . Consequently, the 5 mL of 20 mM H_2O_2 was used in investigating the effect of other parameters.

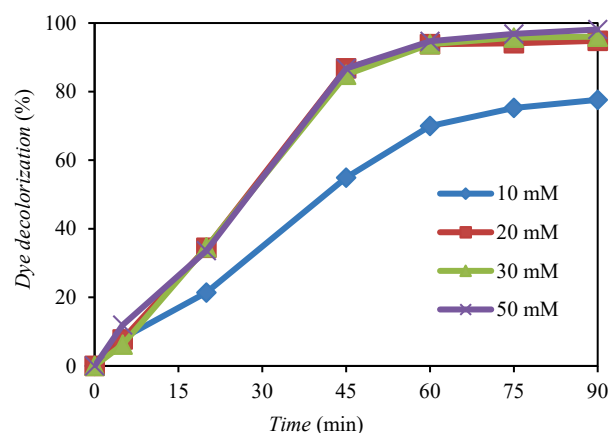


Figure 7 Effect of the H_2O_2 concentration on decolorization ($C_0 = 100 \text{ mg/L}$, $X_0 = 1 \text{ g/L}$, $\text{pH} = \text{natural pH} (\approx 5.2)$ of dye solution, $T = 298 \text{ K}$)

3.3.2 Effect of the Initial pH

One of the desired aims of heterogeneous Fenton-like reactions is developing new heterogeneous catalysts to extend the pH range. The pH is a decisive parameter which influences the efficiency of the Fenton-like process since pH has an important impact on the stability of H_2O_2 , in controlling the catalytic activity, and the dominant iron species [16].

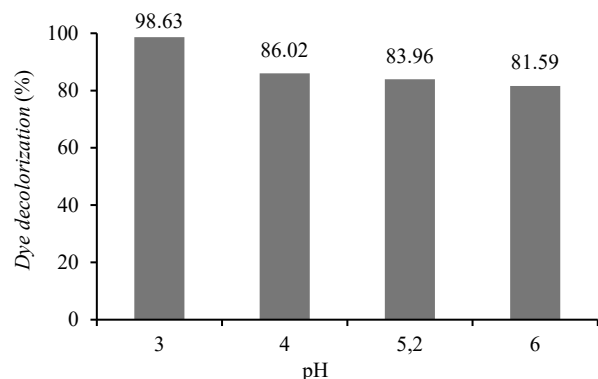


Figure 8 Effect of initial pH on decolorization ($C_0 = 100 \text{ mg/L}$, $X_0 = 1 \text{ g/L}$, $C_{\text{H}_2\text{O}_2} = 5 \text{ mL of } 20 \text{ mM}$, $T = 298 \text{ K}$, $t = 45 \text{ min}$)

The effect of the initial pH of the solution on the catalytic Fenton like decolorization of PR MX-5B was researched at the pH values of 3.0, 4.0, natural pH (≈ 5.2), and 6.0. As seen from Fig. 8, the highest decolorization percentage (98.63%) was obtained at pH 3.0. Moreover, remarkably higher and close decolorization yields (nearly 86%, 84%, and 82%) were achieved at the pH values of 4.0, 5.2, and 6.0, respectively. It was reported in the literature that the reaction between iron-containing minerals and H_2O_2 could sufficiently degrade the organic molecules at higher pH values than at acidic pHs [17]. Furthermore, the formation of $\text{Fe}(\text{OH})_3$ can be prevented due to the ionic cross-links between the Fe ions and the carboxyl group on alginate chains. Therefore, working at natural pH values of dyestuff could avoid the formation of iron sludge at higher pH values.

These outcomes showed that the Fe-Alg gel beads/ H_2O_2

catalytic system could effectively continue in the catalytic degradation of RR MX-5B at the pH range of 3 and 6, indicating the apparent improvement in pH-tolerance.

3.3.3 Effect of Catalyst Concentration

The effect of catalyst concentration on the PR MX-5B decolorization was presented in Fig. 9.

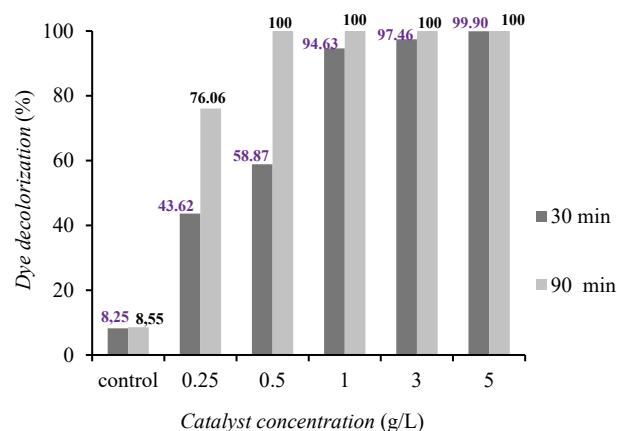


Figure 9 Effect of Fe-Alg gel beads concentration on decolorization ($C_0 = 100 \text{ mg/L}$, $C_{\text{H}_2\text{O}_2} = 5 \text{ mL of } 20 \text{ mM}$, $\text{pH} = 3.0$, $T = 298 \text{ K}$, control represents the dye solution without an added catalyst)

Five different catalyst concentrations, from 0.25 g/L to 5.0 g/L, were selected to determine dye decolorization efficiencies. Besides that, a control experiment was carried out without an added catalyst by using the dye solution containing only the H_2O_2 . As seen in Fig. 9, the control (free catalyst concentration) showed nearly 8.25% PR MX-5B dye decolorization efficiency, and decolorization percentages were increased with an increment in catalyst concentration from 0.25 g/L to 5 g/L. This result arises from the increment in the number of active sites of Fe-Alg gel beads. Therefore, H_2O_2 decomposition improves and leads to the formation of more OH^\bullet radicals.

Moreover, when the Fe-Alg gel beads' concentrations were 0.5, 1.0, 3.0, and 5.0 g/L, $\approx 100\%$ dye decolorization percentages were obtained for 100 mg/L PR MX-5B concentrations at the end of the 90 min reaction time. Moreover, when catalyst concentration increased, the reaction time needed to obtain the 100% decolorization yield decreased. Hence, the 5 g/L of catalyst concentration was chosen as the optimum one to achieve the decolorization removal of high concentrations of dye solutions in a short time.

3.3.4 Effect of Initial Dye Concentration

The effect of initial contaminant concentration on decolorization was shown in Fig. 10. In this study, nearly 100%, 100%, and 96%, 77%, 76%, and 71% decolorization yields were achieved for the initial dye concentrations from 50 mg/L to 500 mg/L, at the end of the 180 min reaction time. Namely, all of the dyestuffs were decolorized in low

concentrations of PR MX-5B, but lower decolorization yields were obtained in high initial dye concentrations.

This phenomenon may be due to the concentration of PR MX-5B affecting the generation of OH^\bullet radicals and their scavengers. Less scavenging of hydroxyl radicals is formed

when the initial organic pollutant concentration was increased without an overlap in its optimum value [7].

Therefore, initial dye concentrations of 50 mg/L, 100 mg/L, and 200 mg/L were selected as suitable for the optimal period needed for the reaction which will prevent the acceleration of the formation of radical OH^\bullet scavengers.

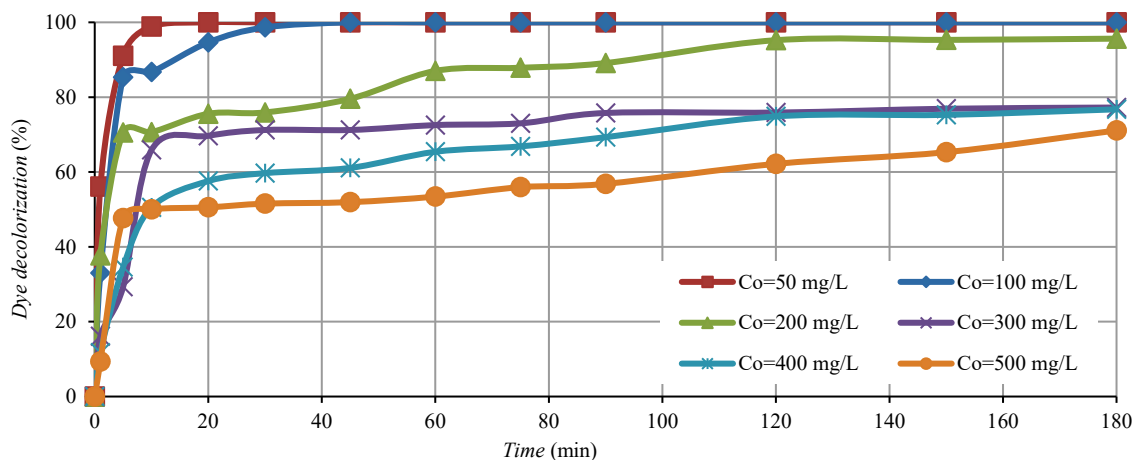


Figure 10 Effect of the initial dye concentration on decolorization ($X_0 = 5$ g/L, $C_{\text{H}_2\text{O}_2} = 5$ mL of 20 mM, $\text{pH} = 3.0$, $T = 298$ K)

Moreover, many studies were carried out in literature for the removal of PR MX-5B by Fenton processes. Some examples of these studies were presented in Tab. 2. In this study, comparing the studies given in Tab. 2, high PR MX-5B decolorization efficiencies were obtained in the studied

range of initial dye concentrations. Therefore, the obtained results showed that effective removal efficiencies can be achieved for wastewaters including high concentrations of PR MX-5B in a single step by using Fe-Alg gel beads as a heterogeneous Fenton-like catalyst.

Table 2 Some of the examples on the studies for the removal of PR MX-5B from the literature

Catalyst	Method	$C_{[\text{PR MX-5B}]}$	D (%)	t (min)	Ref.
Ferric sulfate	Fenton-like decolorization	40 ppm	100	60	[18]
TiO_2	Photocatalytic oxidation	40 ppm	100	20	[19]
Iron(III) Nitrate Nonahydrate and Iron(II) Sulfate Heptahydrate	Fenton ($\text{H}_2\text{O}_2/\text{Fe}^{2+}$) and Fenton-like degradation ($\text{H}_2\text{O}_2/\text{Fe}^{3+}$)	0.1 mM	> 99	480	[20]
Fe/Walnut shell based Activated carbon	Photo Fenton-like process	100 ppm	88.3	120	[21]
Fe/K-Activated carbon	Fenton-like process	100 ppm	91.6	120	[22]
Fe-Alg gel beads	Fenton-like decolorization	50 – 500 ppm	100 – 71	180	This study

3.4 Kinetic Study

Zero, first and second order reaction kinetics were used in order to determine the kinetics of the Fenton-like decolorization of PR MX-5B. The zero, first, and second order rate equations are written as below (Eqs. (11) – (13)) [23]:

$$0^{\text{th}} \text{ order reaction kinetics: } \frac{dC}{dt} = -k_0 \quad (11)$$

$$1^{\text{st}} \text{ order reaction kinetics: } \frac{dC}{dt} = -k_1 C \quad (12)$$

$$2^{\text{nd}} \text{ order reaction kinetics: } \frac{dC}{dt} = -k_2 C^2 \quad (13)$$

Where: C is the concentration of the dye solution; k_0 , k_1 , k_2 - apparent kinetic rate constants of zero, first, and second order reaction kinetics, respectively; t is the reaction time.

When integrating the Eqs. (10) – (13), the following equations are obtained (Eqs. (14) – (16)):

$$C_t = C_0 - k_0 t \quad (14)$$

$$C_t = C_0 e^{-k_1 t} \quad (15)$$

$$\frac{1}{C_t} = \frac{1}{C_0} + k_2 t \quad (16)$$

Where C_t is the concentration of PR MX-5B at any reaction in the time t .

The reaction kinetics for the decolorization with the regression coefficients was calculated. According to the results, the decolorization kinetics of PR MX-5B fitted the 2nd order kinetics well when comparing the regression coefficients (R^2) which were determined from the reaction kinetics. The regression value of the 2nd order reaction kinetics was 0.995, which was much higher than the 0th order ($R^2 = 0.4611$) and the 1st order reaction kinetics ($R^2 = 0.5631$),

(data not shown). The second order reaction kinetics for the decolorization of PR MX-5B was presented in Fig. 11 for this study.

Therefore, the decolorization of PR MX-5B best fitted the second order reaction kinetics of the type:

$$-r_{[\text{PR MX-5B}]} \approx k_2 C_{\text{PR MX-5B}}^2$$

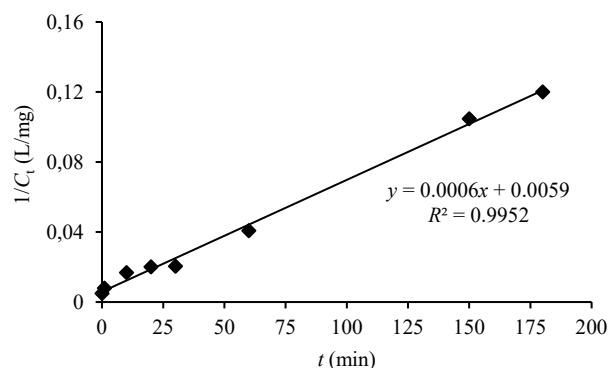


Figure 11 The 2nd order reaction kinetics for the decolorization of PR MX-5B. Experimental conditions: $C_0 = 200$ mg/L; $\text{CH}_3\text{CO}_2\text{H} = 5$ mL of 20 mM, pH = 3.0, $X_0 = 5$ g/L, $T = 298$ K, $t = 180$ min

As seen from Fig. 11, the apparent kinetic rate constant, k_2 was determined to be 0.0006 L/mg.min at the reaction conditions of PR MX-5B concentration: 200 mg/L; H_2O_2 concentration: 20 mM; initial pH of solution: 3.0; catalyst concentration: 5 g/L; temperature 298 K and 180 min reaction time.

4 CONCLUSION

In the presented study, iron-alginate gel beads were synthesized and their utilization was investigated as a heterogeneous Fenton-like catalyst for the decolorization of an azo dye, PR MX-5B. The breakdown of the azo band in the chromophore of the dye with OH^\bullet radicals was demonstrated by the UV-vis by spectral analysis, and the reaction time due to the absorption peaks declined as time passed and finally disappeared.

The Fenton-like degradation was fairly efficient, and high dye decolorization percentages from 100% to 71% were obtained in the range of 50 – 500 mg/L dye concentrations, at the end of the 180 min reaction time. Furthermore, the kinetic study demonstrated that the reaction kinetics of PR MX-5B fitted the second order kinetics well.

Moreover, high decolorization percentages were obtained in the range of the initial pH of 3.0– 6.0, showing the pH tolerance of the system. The results indicated that Fe-alginate gel beads could be used as a catalyst effectively in the heterogeneous Fenton-like reactions for the removal of azo dyestuffs. The using of Fe-Alg gel beads offers several advantages owing to their ease of production, high capability of decomposing the hydrogen peroxide and formation the OH^\bullet radicals to provide effective decolorization. In addition, they can be easily separated from the reaction medium after the procedure and overcome the formation of iron sludge by

classical Fenton decomposition Fenton's degradation, especially at higher pH values than at acidic values.

5 REFERENCES

- [1] Kasiri, M. B., Aleboyeh, H., & Aleboyeh, A. (2008). Degradation of Acid Blue 74 using Fe-ZSM5 zeolite as a heterogeneous photo-Fenton catalyst. *Applied Catalysis B: Environmental*, 84(1-2), 9-15. <https://doi.org/10.1016/j.apcatb.2008.02.024>
- [2] Pokharia, A. & Ahluwalia, S. S. (2016). Biodecolorization and degradation of xenobiotic azo dye-Basic Red 46 by *Staphylococcus epidermidis* MTCC 10623. *Int. J. Res. Biosciences*, 5(2), 10-23.
- [3] Nidheesh, P. V., Gandhimathi, R., & Ramesh, S. T. (2013). Degradation of dyes from aqueous solution by Fenton processes: a review. *Environmental Science and Pollution Research*, 20(4), 2099-2132. <https://doi.org/10.1007/s11356-012-1385-z>
- [4] Guivarch, E., Trevin, S., Lahitte, C., & Oturan, M. A. (2003). Degradation of azo dyes in water by electro-Fenton process. *Environmental Chemistry Letters*, 1(1), 38-44. <https://doi.org/10.1007/s10311-002-0017-0>
- [5] Wang, N., Zheng, T., Zhang, G., & Wang, P. (2016). A review on Fenton-like processes for organic wastewater treatment. *Journal of Environmental Chemical Engineering*, 4(1), 762-787. <https://doi.org/10.1016/j.jece.2015.12.016>
- [6] Mazilu, I., Ciotonea, C., Chiriac, A., Dragoi, B., Catrinescu, C., Ungureanu, A., S. Petit, S. Royer, and Dumitriu, E. (2017). Synthesis of highly dispersed iron species within mesoporous (Al)- SBA-15 silica as efficient heterogeneous Fenton-type catalysts. *Microporous and Mesoporous Materials*, 241, 326-337. <https://doi.org/10.1016/j.micromeso.2016.12.024>
- [7] Li, R., Gao, Y., Jin, X., Chen, Z., Megharaj, M., & Naidu, R. (2015). Fenton-like oxidation of 2, 4-DCP in aqueous solution using iron-based nanoparticles as the heterogeneous catalyst. *Journal of colloid and interface science*, 438, 87-93. <https://doi.org/10.1016/j.jcis.2014.09.082>
- [8] Kuang, Y., Wang, Q., Chen, Z., Megharaj, M., & Naidu, R. (2013). Heterogeneous Fenton-like oxidation of monochlorobenzene using green synthesis of iron nanoparticles. *Journal of colloid and interface science*, 410, 67-73. <https://doi.org/10.1016/j.jcis.2013.08.020>
- [9] Ren, H., Gao, Z., Wu, D., Jiang, J., Sun, Y., & Luo, C. (2016). Efficient Pb (II) removal using sodium alginate-carboxymethyl cellulose gel beads: preparation, characterization, and adsorption mechanism. *Carbohydrate polymers*, 137, 402-409. <https://doi.org/10.1016/j.carbpol.2015.11.002>
- [10] Rosales, E., Iglesias, O., Pazos, M., & Sanromán, M. A. (2012). Decolorization of dyes under electro-Fenton process using Fe alginate gel beads. *Journal of hazardous materials*, 213, 369-377. <https://doi.org/10.1016/j.jhazmat.2012.02.005>
- [11] Lin, Y. C. & Lee, H. S. (2010). Effects of TiO_2 coating dosage and operational parameters on a TiO_2/Ag photocatalysis system for decolorization Procion red MX-5B. *Journal of hazardous materials*, 179(1-3), 462-470. <https://doi.org/10.1016/j.jhazmat.2010.03.026>
- [12] Kannan, A. & Thambidurai, S. (2008). Removal of hexavalent chromium from aqueous solution using activated carbon derived from palmyra palm fruit seed. *Bulletin of the Chemical Society of Ethiopia*, 22(2). <https://doi.org/10.4314/bcse.v22i2.61282>
- [13] Li, P., Dai, Y. N., Zhang, J. P., Wang, A. Q., & Wei, Q. (2008). Chitosan-alginate nanoparticles as a novel drug delivery

- system for nifedipine. *International journal of biomedical science: IJBS*, 4(3), 221.
- [14] Cihanoğlu, A., Gündüz, G., & Dükkancı, M. (2015). Degradation of acetic acid by heterogeneous Fenton-like oxidation over iron-containing ZSM-5 zeolites. *Applied Catalysis B: Environmental*, 165, 687-699. <https://doi.org/10.1016/j.apcatb.2014.10.073>
- [15] Zhu, G., Yin, J., Wan, P., Zheng, H., & Deng, B. (2015). Fe (III)/H₂O₂-Like System for Removal of Azo Dye from Aqueous Solution. *Separation Science and Technology*, 50(14), 2220-2229. <https://doi.org/10.1080/01496395.2015.1020551>
- [16] Xu, L. & Wang, J. (2011). A heterogeneous Fenton-like system with nanoparticulate zero-valent iron for removal of 4-chloro-3-methyl phenol. *Journal of hazardous materials*, 186(1), 256-264. <https://doi.org/10.1016/j.jhazmat.2010.10.116>
- [17] Xue, X., Hanna, K., & Deng, N. (2009). Fenton-like oxidation of Rhodamine B in the presence of two types of iron (II, III) oxide. *Journal of hazardous materials*, 166(1), 407-414. <https://doi.org/10.1016/j.jhazmat.2008.11.089>
- [18] Wu, C. H. (2008). Decolorization of CI Reactive Red 2 in O₃, Fenton-like and O₃/Fenton-like hybrid systems. *Dyes and Pigments*, 77(1), 24-30. <https://doi.org/10.1016/j.dyepig.2007.03.002>
- [19] So, C. M., Cheng, M. Y., Yu, J. C., & Wong, P. K. (2002). Degradation of azo dye Procion Red MX-5B by photocatalytic oxidation. *Chemosphere*, 46(6), 905-912. [https://doi.org/10.1016/S0045-6535\(01\)00153-9](https://doi.org/10.1016/S0045-6535(01)00153-9)
- [20] Hsueh, C. L., Huang, Y. H., Wang, C. C., & Chen, C. Y. (2005). Degradation of azo dyes using low iron concentration of Fenton and Fenton-like system. *Chemosphere*, 58(10), 1409-1414. <https://doi.org/10.1016/j.chemosphere.2004.09.091>
- [21] Palas, B., Ersöz, G., & Atalay, S. (2016). Heterogeneous photo Fenton-like oxidation of Procion Red MX-5B using walnut shell based green catalysts. *Journal of Photochemistry and Photobiology A: Chemistry*, 324, 165-174. <https://doi.org/10.1016/j.jphotochem.2016.03.031>
- [22] Palas, B., Ersöz, G., & Atalay, S. (2017). Green catalysts for Fenton-like oxidation of Procion Red MX-5B: Influence of the activation method and the reaction parameters on dye removal. *Separation Science and Technology*, 52(3), 404-420. <https://doi.org/10.1080/01496395.2016.1252399>
- [23] Sun, S. P., Li, C. J., Sun, J. H., Shi, S. H., Fan, M. H., & Zhou, Q. (2009). Decolorization of an azo dye Orange G in aqueous solution by Fenton oxidation process: Effect of system parameters and kinetic study. *Journal of hazardous materials*, 161(2-3), 1052-1057. <https://doi.org/10.1016/j.jhazmat.2008.04.080>

Authors' contacts:

Memduha ERGÜT, PhD, Research Assistant

Corresponding author

Mersin University, Faculty of Engineering, Department of Chemical Engineering

Mersin University Çiftlikköy Campus F Block, 33343, Yenişehir/Mersin, Turkey

Tel./Fax, +90 (0324) 361 00 01-17371 / +90 (0324) 361 00 32

E-mail: memduha.ergut@gmail.com

Ayla ÖZER, PhD, Professor

Mersin University, Faculty of Engineering, Department of Chemical Engineering

Mersin University Çiftlikköy Campus F Block, 33343, Yenişehir/Mersin, Turkey

Tel./Fax, +90 (0324) 361 00 01-17370 / +90 (0324) 361 00 32

E-mail: ayozer4@gmail.com

THE INFLUENCE OF PRE-TREATMENT BY PRIMING ON THE CMY REPRODUCTION QUALITY PRINTED WITH ELECTROINK

Marko MORIĆ, Igor MAJNARIĆ, Klaudio PAP, Slaven MILOŠ

Abstract: In order to achieve high-quality colour prints by using an electrophotography printing machine, it is necessary to prepare the printing substrate. Additional treatment may be needed to achieve the better substrate surface tension for optimally attaching the liquid electrophotography ink to the fine art paper. To ensure good adhesion of the ink to the printing substrate, a pre-treatment by application of surface tension regulating fluids (priming) may be used. In this paper, we will determine the possibility of realization of the CMY prints and the possibility of achieving an increased range of tone value in electrophotography digital printing. Based on the results, the new possible values of reproduced CMY tones will be determined so that a higher print quality can be achieved. The application of primers changes the optical properties of the paper printing substrate, which directly affects the colour reproduction. An additional 100 % increase of primer application will result in total colour changes of $\Delta E_C (0,5g/m^2 \text{ \& } 1g/m^2) = 0,06$, $\Delta E_M (0,5g/m^2 \text{ \& } 1g/m^2) = 0,42$, $\Delta E_Y (0,5g/m^2 \text{ \& } 1g/m^2) = 0,08$. In this scenario, 1 g/m² application of primer will produce the most intense full tones and therefore the highest quality prints for all three ElectroInk process colours.

Keywords: CIELab; colour difference; digital colour offset; ElectroInk; in-line primer; pre-treatment

1 INTRODUCTION

In order to achieve uniform and well-defined surface properties, it is necessary to treat the surface with certain substances. Most commonly these are acrylic polymers, vinyl ether polymers, polystyrene butadiene, polyesters, alkyd resins, phenolic resins, vulcanized thermoplastic elastomers, olefin thermoplastic elastomers, EVOH (ethyl vinyl alcohol copolymer) resins, polyvinyl chloride, resin oils, waxes [1]. Their application gives the possibility of achieving better mechanical properties of the materials, as well as their long-term use.

Depending on the aggregate condition of the printing media, the production speed and the thickness of the coating, different coating methods are applied - coating performed by pouring, direct immersion of the material in the coating, application through a narrow crack (extrusion), application with chamber squeegees, application by Mayer rod. All of these procedures are based on the principle of uniform application with 100 % coverage of the surface. Differences between these techniques are evident in production productivity, which can range from 50 to 1.500 m/min [2]. In order to achieve such production speeds, the viscosity of the coating fluid, whose value can vary from 10 to 20.000 mPa·s, is crucial. Thus, only ideal fluids (subject to Newton's laws) will achieve maximum coating productivity. The realized coating thicknesses can be different, depending on the coating techniques, ranging from 3 to 300 g/m² [3].

Coating techniques can also be applied to standard printing techniques such as flexographic printing, offset printing and contactless inkjet printing. The advantage of such coating technologies is the high productivity with the possibility of partial coating (due to the use of classic printing forms).

2 THEORETICAL PART

The in-line unit used with the HP Indigo WS 6800 series machines has two basic components. The coating is first

charged with the substrate, followed by the application of a surface tension regulating fluid (primer). The printing substrate refining unit is used to apply a specific fluid that will affect the surface tension of the printing substrate and thus to the acceptance of ElectroInk [4]. Fig. 1 shows the design of the pre-treatment unit on HP Indigo printing machines.

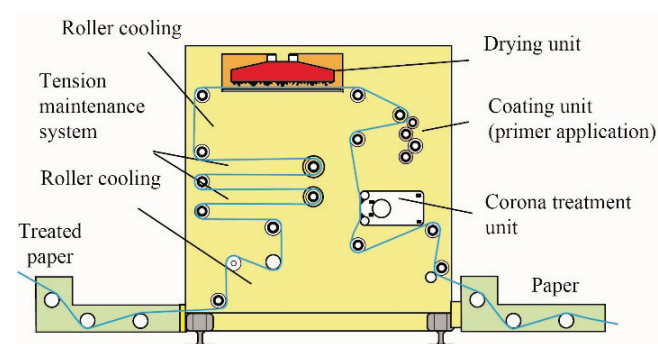


Figure 1 HP Indigo unit for pre-treatment of printing substrates.

The printing substrate refining unit works by unwinding the roll with the printing pad and passing it through the corona unit. Therefore, the surface of the printing substrate is charged with ionized air in order to create a microstructure (roughness) that will improve the ink application on the printing substrate.

After that, the roll with the printing substrate enters a central coating unit containing five rollers. This results in a uniform application of the primer to the surface of the printing substrate. The amount of coating applied depends on the diameter of the primer coating roller, for example a coating thickness of 74 mm achieves a coating of 0,5 g/m², while a coating diameter of 84 mm achieves a coating of 1 g/m². In order for the primer to dry, the roll with the printing substrate must pass through a warm air drying tunnel and then reach the cooling rollers. Such printing substrate at the

end of the refining unit enters the printing unit in HP Indigo where it is imprinted on a previously applied primer.

The priming unit, with the exception of the primer form roller, contains a pressure roller, a primer transfers roller, a distributor roller and a primer fountain roller (Fig. 2). The primer fountain roller is used to apply the ink from the ink tank, the distribution roller spreads the primer over the entire surface of the print, and the primer transfer roller connects them. The highest pressure roller presses the roller with the printing substrate in a controlled manner to form the high-quality primer coating on the printing substrate. It is this construction that ensures uniform treatment and prints without any visible defects.

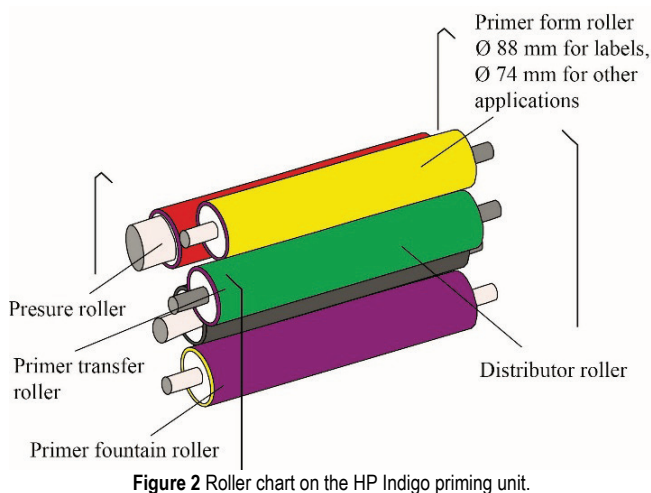


Figure 2 Roller chart on the HP Indigo priming unit.

2.1 Digital Colour Offset

Digital colour offset is a specially developed electrophotography technique that utilizes the liquid ElectroInk. To generate a single imprint, the machine should contain a three cylinders' satellite unit (PIP cylinder, Intermediate cylinder and Impression cylinder), which serves for indirect transfer of ElectroInk to the printing substrate (Fig. 3).

ElectroInk is a special liquid electrophotography ink that, depending on the colour tone, contains the following components: volatile ISOPAR oil (90 % to 95 %), pigment particles (5 % to 8 %) and charge regulator - Image agent (1 % to 3 %). The liquid ink is prepared immediately prior to printing, using densitometers (the proportion of dry matter content in liquid) whose value is 1.70 g/cm³ [5], [6]. The selective behaviour of ElectroInk depends on the established electrostatic field between the developer roller and the PIP cylinder or the PIP cylinder and the Intermediate Transfer Drum [7], [8].

The Liquid ElectroInk is transferred to the heated surface of the Intermediate Transfer Drum which has a soft offset rubber on its surface. During that process, ElectroInk changes its aggregate state and turns into a sticky paste [9]. In order to achieve a complete transfer, the offset rubber cover must meet two basic properties: a) it must be conductive; b) it must be heated to a temperature above 100° [10].

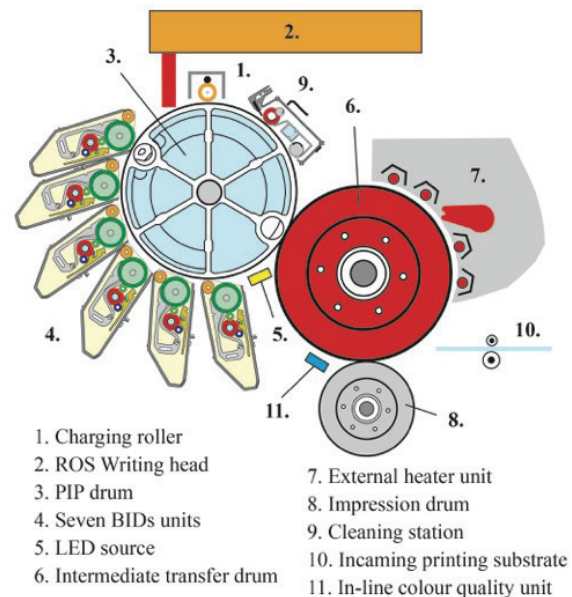


Figure 3 Schematic representation of printing process on digital colour offset (HP Indigo WS 6800 machine).

In order to ensure the ability to print in photo quality, BIDs should be added consist CMY colours. A parameter that significantly influences the quality is the adequate print of raster elements that can be directly controlled by activating laser power and by adjusting the colour transfer by changing the voltage on BID units and by selecting the appropriate printing substrate [11], [12], [13].

3 EXPERIMENTAL PART

For experimental printing, an HP Indigo WS 6800 was used with an integrated in-line primer injector and a corona unit [14]. The RIP (Raster Image Processor) used for separation of colour extracts was the ESCO HPE ProLiant ML350, while using a standard calibration curve LUT 0.5 and 180 lpi raster lines. For testing, the standard print form "FOGRA Image Quality" of the SRA3 dimension was used, out of which we applied Cyan (C), Magenta (M) and Yellow (Y) Separation [15].

Table 1 Characteristics of used fine art paper.

Paper features - Condat digital gloss RL			
Attribute	Unit of measure	Value	Tolerance
Grammage	g/m ²	90	±4%
Thickness	µm	70	±4%
Opacity	%	89	-2.0
Whiteness D65/10°	%	118.7	±3
Gloss of paper	%	70	±5
Smoothness by Bekk	s	1400	±30%

The paper substrate used for the experiment was 90 g/m² Condat digital gloss RL (gloss coated fine art paper). The composition of such paper contains: 50% virgin cellulose, 40% calcium carbonate, 4.5% water and 5.5% latex and sizing agents [16]. Other properties of used fine art paper are shown in Tab. 1.

During the printing process paper was treated by means of two different primer depositions (Michem® In-Line

Primer 030 manufactured by Michelman) in an amount of 0.5 g/m² and 1 g/m² (power of corona was 0 W). Michem® In-Line Primer 030 is a special substance manufactured for the in-line coating of HP Indigo printing machines, series WS 6800. This primer improves the adhesion of ElectroInk and enables better transfer of colour to a low surface energy substrate [17]. The technical characteristics of this primer are shown in Tab. 2.

Table 2 Characteristics of Michem® In-Line Primer 030

Technical characteristics of in-line primer	
Characteristic	Value
Appearance	White fluid
pH	8.0 - 10.0
Viscosity (cps)	<100
Application Number	1
Application rate	60 rpm
Manufacturer	USA, Belgium, Singapore

Since the measurements were conducted in industrial conditions, a surface tension measurement method was applied in accordance with the norm ASTM D-2578, with the use of indicator fluids containing 2-ethoxyethanol & formamide [18]. In this way, we have avoided a greater pause between examination and testing phases. DYNE test results of surface tension without primer and after priming is given in Tab. 3.

Table 3 Surface tension measured by the DYNE method according to ASTM D-2578

Position of measure	Primer 0 g/m ²	Primer 0,5 g/m ²	Primer 1 g/m ²
1	36	34	38
2	36	34	38
3	36	34	38
4	36	34	38
5	36	34	38
6	36	34	38
7	36	34	38
Average	36	34	38

Table 4 Climatic conditions in the production plant during the measurement

Samples	Temperature °C	Relative humidity %
Paper – primer 0 g/m ²	24,2	49
Paper – primer 0,5 g/m ²	24,7	51
Paper – primer 1 g/m ²	24,9	52
Min.	24,2	49
Max.	24,9	52
Average	24,60	50,67

Characterization of the printing surface is possible by determining the surface tension. Due to the specificity of testing, a DYNE test with indicator fluids of 34 to 44 mN/m (DYNE/cm) was applied in the experiment. This method was chosen because of the precision of sample measurements taken immediately after the printing substrate was released from the machine. Room temperature and relative humidity are very important for this method. Tab. 4 shows the climatic conditions used in the measurement of surface tension of the treated materials.

To determine the CMY reproducibility, the 4-generation X-rite eXact [19] colorimeter and spectrophotometer was

used, while the Personal IAS was used for visual analysis of magnified CMY raster elements on the print [20]. The obtained CIE $L^*a^*b^*$ and CIE LAB ΔE results were compared to each other with the Origin 8.0 software. During experimental printing, the different power values of the primer were used (0 g/m², 0,5 g/m² and 1 g/m²).

As a result, nine different samples were produced during the experimental printing, as shown in the chronological experiment (Fig. 4).

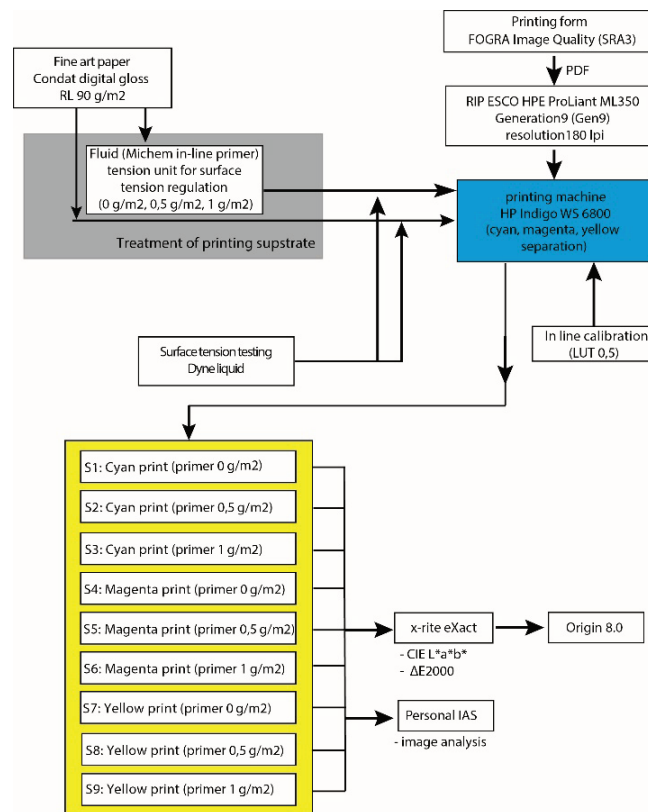


Figure 4 Schematic representation of the chronological experiment

4 RESULTS AND DISCUSSION

When printing on demanding printing substrates (characterized by low surface absorbency), it is necessary to treat the fluids that regulate the surface tension. The results are better bonding of ElectroInk as well as the optical properties of the printing substrate.

The colour change $\Delta E < 1$ represents visually invisible colour changes, ΔE 1 to 3 colour range is the standard, ΔE 3 to 5 is the acceptable colour deviation range and $\Delta E > 5$ is the unacceptable deviation area. Figs. 5, 6 and 7 show colour changes for process colours (cyan, magenta and yellow).

The reproduction curves created by applying a different amount of primer completely realize colour changes (ΔE), which is true for all experimental samples tested. This means that visible colour changes will be achieved by priming the printing substrate (paper) without applying the corona treatment.

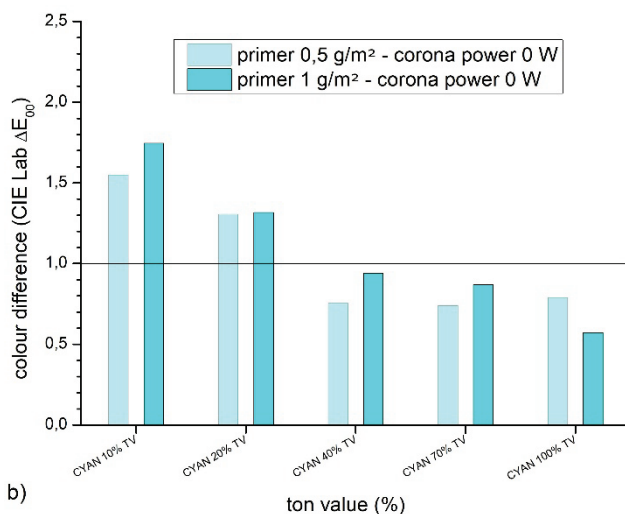
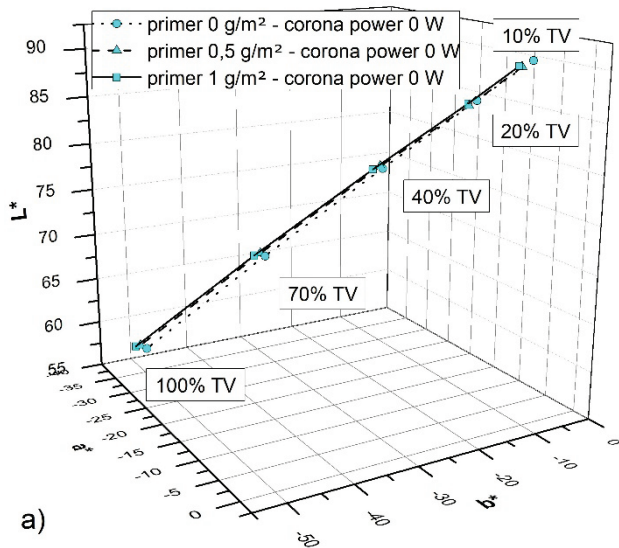


Figure 5 Colour changes of cyan prints: a) the primer variation (0 g/m², 0.5 g/m² and 1 g/m²) without corona power (0 W); b) ΔE colour deviations

On untreated paper without activated corona treatment, greater colour changes will be achieved by applying a primer of 1 g/m² ($\Delta E_{avr.0W} = 1.09$). In this case, the smallest surface coverage analysed will produce a larger colour change than full tone ($\Delta E_{C10\%_0W} = 1.75$; $\Delta E_{C100\%_0W} = 0.57$). Imprints with a smaller amount of primer (0.5 g/m²) produce less colour changes $\Delta E_{avr.0W} = 1.03$. They are also expressed in the brighter areas ($\Delta E_{C10\%_0W} = 1.55$) to stabilize with increasing surface coverage of the tone deviation ($\Delta E_{C100\%_C70\%_40\%_0W} = 0.76$). The application of a primer of 1 g/m² resulted in the deviation of cyan tones, which were expressed more in chromaticity (a^* and b^* coordinates) and less in luminance (L^*). By reducing the amount of primers, cyan prints become more unstable, meaning that in bright areas changes are more pronounced in chromaticity and in higher tones (from 40% to 100% of TV) in brightness. In order to achieve the most saturated cyan print, it is necessary to apply a primer in the amount of 1 g/m² without corona power (0 W).

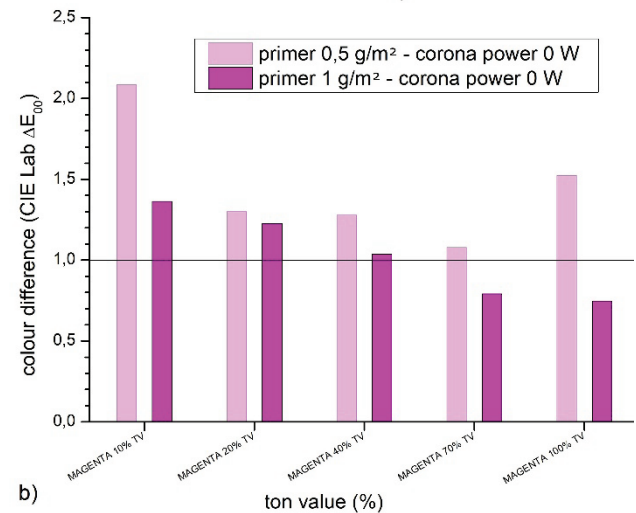
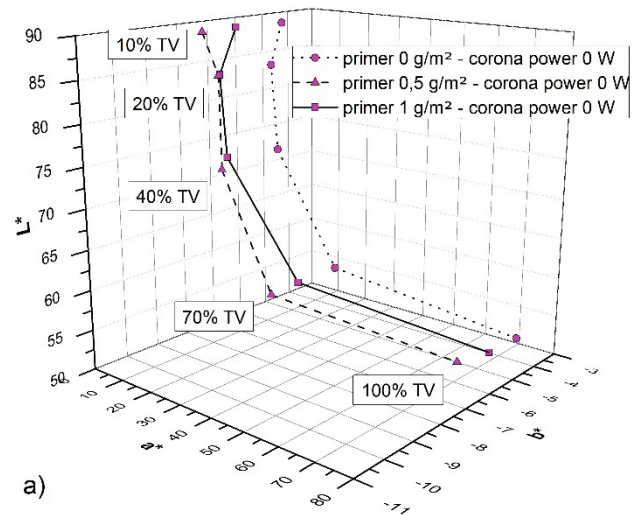


Figure 6 Colour changes of magenta prints: a) the primer variation (0 g/m², 0.5 g/m² and 1 g/m²) without corona power (0 W); b) ΔE colour deviations

Magenta ElectroInk printing produces more pronounced colour changes in case of priming. Without activation of corona power and application of a primer of 0.5 g/m², a clear colour difference is achieved in all measurement fields ($\Delta E_{avr.0W} = 1.45$). Thus, two areas stand out with the values: 10% of TV ($\Delta E_{M10\%_0W} = 2.09$) and 70% of TV ($\Delta E_{M10\%_0W} = 1.08$). With the further increase of primers, the print becomes more stable and achieves an average colour change of $\Delta E_{avr.0W} = 1.03$. On such prints, larger surface coverage achieves a smaller colour change ($\Delta E_{M100\%_M70\%_0W} = 0.76$) and the smallest largest ($\Delta E_{M10\%_0W} = 1.36$). By applying a fluid to refine the printing substrate, magenta experimental prints will retain a specific reproduction curve that changes up to 70% of TV in brightness and then sharply shifts to the coordinate (a^*). However, priming without activating corona power produces the largest colour changes that are expressed along the chromaticity axes. The best full tone reproduction will be achieved in scenarios without corona treatment and with primer application (1 g/m²).

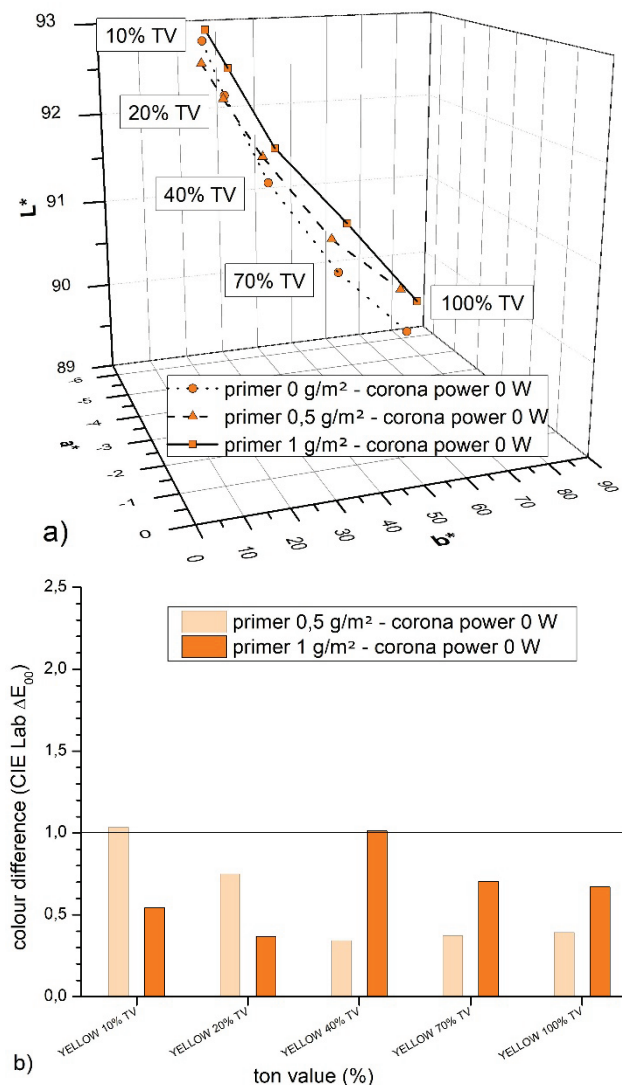


Figure 7 Colour changes of yellow prints: a) the primer variation (0 g/m², 0.5 g/m² and 1 g/m²) without corona power (0 W); b) ΔE colour deviations

Of all the analysed colour separations of yellow tones, by the priming process, it showed the least colour deviation and maximum uniformity of tones within the reproduction curve. Such prints, regardless of the applied corona power, will not produce larger colour changes than $\Delta E > 1$. Experimental application of primer in the amount of 0.5 g/m² and without corona treatment will result in an average colour change of $\Delta E_{avr,0W} = 0.58$. Higher priming (1 g/m²) results in a slight increase in the average yellow colour deviations of $\Delta E_{avr,0W} = 0.66$. The obtained results are characterized by a trend of realized tonal changes where a smaller amount of primers will produce larger deviations in the brighter parts of the image, while a larger amount of primers will produce larger deviations in the darker portions of the image. Only the primed yellow prints (Lab curve with no corona activated) produce a characteristic change that is evident in the rapid growth of chromaticity in the higher tone range. The most pronounced variable is (+ b*). Thus, the best print will be realized with a coating of 1 g/m². The highest-quality yellow colour is created by priming 1 g/m² without activating corona

power. This is due to the chemical composition of the primer, which is not a colourless but light yellow colour. This is also the reason for such results.

In order to visualize realized ΔE changes of CMY tones, experimental prints have been digitized and magnified 60 times (Fig. 8).

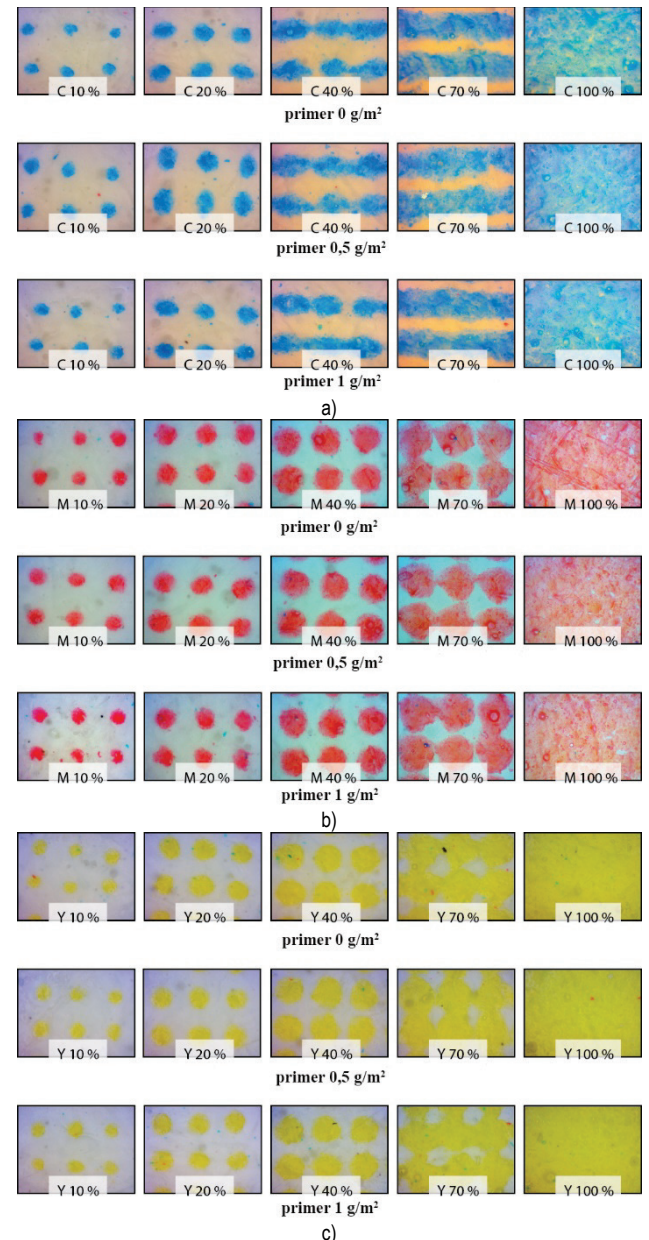


Figure 8 Magnified HP Indigo prints by variations of priming (0 g/m², 0.5 g/m² and 1 g/m²) without corona power (0 W): a) cyan; b) magenta; c) yellow.

Cyan separation is characterized by the connection of raster print elements in the middle and higher tonal values. Such deformation is least visible in primed paper (1 g/m²) which also gives the best print result. Magenta colour separation achieves the most correct raster elements if the printing substrate is primed also with a primer of 1 g/m². Such elements have geometric regularity and the highest print intensity (they have the highest optical density). The yellow separation is also best reproduced with a primer of 1

g/m². With the good application of yellow ElectroInk quality, increase an additional yellowness printing substrate. Yellow prints are much better reproduced than other process separations.

5 CONCLUSION

The aim of this research was to define the effect of refinement of the printing substrate by the method of application of fluid to regulate the surface tension of the printing substrate. Thus treated print media for the purposes of the complete reproduction process require the redefinition of the qualitative properties of digital multicolour reproduction (CIE $L^* a^* b^*$, ton value, imbalance of print) when printing with ElectroInk.

The difference between the two colour changes in cyan prints, due to the variation of the primer (0.5 g/m² and 1 g/m²), is ΔE difference = 0.06. The difference in magenta prints is ΔE difference = 0.42. Regarding the variation of the fluid for regulating the surface tension of the printing substrate, the difference between the two colour changes of the yellow prints is ΔE difference = 0.08.

Thus, the best separation of elements is reproduced with cyan and yellow ElectroInk. The changes made by priming the printing substrate are not visible, but they are important in the whole process of calibrating the electrophotography printing machine. In combination with other parameters of the electrophotography printing process, those changes still affect the final colour results.

6 REFERENCES

- [1] McLeod, D. M. (2006). Wire-Wound Rod Coating. In *Coatings Technology Handbook, Third.*, A. A. Tracton, Ed. London: CRC Press - Taylor & Francis Group, 18.1-18.7. <https://doi.org/10.1201/9781420027327.pt2>
- [2] Bolanča, S., Majnarić, I., & Golubović, K. (2015). *Packaging Printing Today. Acta Graph.* 236, 26(4), 27-33.
- [3] Bolanča, S. (2013). *Tisak ambalaže*. Zagreb, Hrvatska: Hrvatska sveučilišna naklada.
- [4] HP, Operating the Inline Primer. Available: ftp://ftp.hp.com/pub/gsb/Indigo/Industrial/Software/CN-ILP_User_Guide/CA394-11321.pdf.
- [5] Benzion, L. et al. (1988). Toner for Use in Compositions for Developing Latent Electrostatic Images, Method of Making the Same, and Liquid Composition Using the Improved toner. 4,794,651,
- [6] Yaacov, A. (1991). Humidity Tolerant Charge Director Compositions, p. 6.
- [7] Forgacs, P. & Teishev, A. (2013). Electro-Rheological Model of HP Indigo ElectroInk. *J. Imaging Sci. Technol.*, 56(4), 1-9. <https://doi.org/10.2352/J.ImagingSci.Technol.2013.57.3.030403>
- [8] Majnaric, I., Golubovic, K., Bolanca, S., & Modric, D. (2011). Voltage Effect on Developing Process and B&W Reproduction, *DAAAM Int. Sci. B.* 2010. <https://doi.org/10.2507/daaam.scibook.2010.45>
- [9] Majnarić, I., Bolanča, I., Bolanča, Z., & Milković, M. (2005). Conditions in digital printing of packaging on the print quality. In *International Conference on Digital Production Printing and Industrial Applications, DPP2005 - Final Program and Proceedings*, vol. 2005.
- [10] Gazit, A. (1998). Imaging Apparatus and Intermediate Transfer Blanket Therefor, p. 20.
- [11] Majnaric, I., Modric, D., & Golubovic, K. (2008). Influence of laser power output on the quality of colour imaging. In *Annals of DAAAM and Proceedings of the International DAAAM Symposium*.
- [12] Majnaric, I., Bolanca, Z., & Bolanca Mirkovic, I. (2009). The influence of the developer drum voltage on screen dot formation. In *Annals of DAAAM and Proceedings of the International DAAAM Symposium*.
- [13] Majnarić, I., Hladnik, A., Muck, T., & Bolanca Mirkovic, I. (2015). The influence of ink concentration and layer thickness on yellow colour reproduction in liquid electrophotography toner. *Tehnicki vjesnik*, 22(1), 145-150. <https://doi.org/10.17559/TV-20140321230455>
- [14] Hp, T. et al. HP Indigo WS6800 Digital Press and packaging production.
- [15] Kraushaar, A. (2018). Process Standard Digital Handbook Munich: Fogra Research Institute for Media Technologies.
- [16] Lecta group (2018). Condac data sheet, Available: https://cmspro.lecta.com/Digital/LectaProductCertificates/SafetyDataSheet_Condac.pdf.
- [17] Sheet, T. D. (2018, April). Michem ® In-Line Primer 030 Michem ® In-Line Primer 030, 1-2.
- [18] ASTM International, Standard Test Method for Wetting Tension of Polyethylene and Polypropylene Films.
- [19] X-rite (2018). x-rite eXact Spectrophotometer, Available: <https://www.xrite.com/categories/portable-spectrophotometers/exact>.
- [20] QEA, Personal IAS. Available: http://www.qea.com/upload/files/products/datasheet_personalias-newaddr.pdf.

Contact information:

Marko MORIĆ, PhD

(Corresponding author)

University North, Multimedia, Design and Application,
Jurja Križanića 31b, 42000 Varaždin, Croatia
mmoric@unin.hr

Igor MAJNARIĆ, Assoc. Prof.

University of Zagreb, Faculty of Graphic Arts,
Getaldiceva 2, 10000 Zagreb, Croatia
igor.majnarić@grf.hr

Klaudio PAP, Prof.

University of Zagreb, Faculty of Graphic Arts,
Getaldiceva 2, 10000 Zagreb, Croatia
klaudio.pap@grf.hr

Slaven MILOŠ, MSc

Grafiknet Ltd.
Odranska 1/1, 10000 Zagreb, Croatia
slaven.milos@grafiknet.hr

NEAR-INFRARED SPECTROSCOPY AND HIDDEN GRAPHICS APPLIED IN PRINTING SECURITY DOCUMENTS IN THE OFFSET TECHNIQUE

Vilko ŽILJAK, Jana ŽILJAK GRŠIĆ, Denis JUREČIĆ, Tonči JELIČIĆ

Abstract: The method of mixing pigments for the printing technology is expanded with the addition of dual states in security graphics. Near-infrared spectroscopy of colorants used in the offset technique has brought about the concept of targeted design by a selective use of colorants that are visible in the visible region of the electromagnetic spectrum. High-security documents have been observed in the NIR light based on twin colorants which are differentiated according to their NIR absorption in the range from 800 to 1000 nm. New procedures have been developed based on the innovative creation of documents with hidden information. Two graphics are combined with mutating screen elements. A mathematical definition of the new printing screen and an algorithm in PostScript are given. The designer plans and creates graphics, which originate from the twin colorants, by using the INFRAREDESIGN® method. The dual design was investigated using visual and infrared cameras for the forensic area. The components of colorants are determined by the spectroscopy of light absorption with filters in the visual and NIR spectrum.

Keywords: hidden graphics; near-infrared spectroscopy; offset printing; security screen; twin colorants

1 INTRODUCTION

Instrumental observation of near-infrared graphics has initiated the expansion of security printing into the conventional printing practice. Our research into protected coloring is published with applications in digital printing with toners, flexographic printing and with the introduction of image duality in two light spectra [1]. In this paper, the application of twin colorants for the offset technique is published. The twins of colorants are prepared with the corresponding offset colorants from the "scale" – that is, with the set of the C, M, Y, K components for process printing. The INFRAREDESIGN® technology is prepared at high circulation in the newspaper press [2]. The idea of "twin colorants" is partly explained in the work on the application of the inkjet print on textiles for dresses in the clothing industry [3]. This article is based on the INFRAREDESIGN® theory, which gives principles for combining two images [4]. The construction of dual cameras (ZRGB) is described in the article by Ivan Rajković et al. [5], with application in film set design. Cameras are used as an instrument for determining light absorption on pigments and colorants in the industry [6].

This article requires explanations for the abbreviations used in the application of IRD.

The spectrum of the naked eye – red, green, blue (RGB) – is marked as V (visual). Absorption in the segment of the NIR spectrum at 1000 nm is marked as Z. The same abbreviations are used for the corresponding photographic cameras. The hidden image at 1000 nm is called the Z image (Z information). Colorants which do not absorb Z radiation are called V colorants. The observation of colorants through the system of "twins of colors and colorants" [7] has been introduced. The different groups of twin colorants are recognized by the ZRGB camera, which determines and assigns numerical values of the light absorption of colored material [8].

2 RECIPES FOR TWIN COLORS FOR THE OFFSET TECHNIQUE

The numerical values of the recipes for the C, M, Y and K process colorant components are given. The colorants chosen for raster printing are the same as the colors in the security document (Tab. 1). Images of twin colorants at 1000 nm were added using a dual ZRGB camera [8].

Table 1 The recipes of twin colorants with the C, M, Y, K dyes

Colors	RGB (2°)	C, M, Y, K %	C, M, Y, K %	$L^*a^*b^*$
1 Hidden letters, Z	102, 82, 78	65, 69, 67, 0	36, 47, 40, 42	37, 11, 7
2 Dark thin line; V, Z	46, 0, 0	89, 97, 99, 0	38, 85, 78, 65	0, 46, 15
3 Blue-green surface, V	57, 67, 106	89, 71, 42, 0	71, 49, 9, 40	28, 4, -27
4 Gray color, V	129, 134, 151	55, 41, 27, 0	27, 13, 0, 40	56, 0, -11

Twin colorants were photographed with filters at 400, 470, 570, 610 nm and in the visible spectrum. Light absorption was measured by using the Projectina forensic camera [9] for the visible (V – RGB) and near-infrared (NIR – Z) spectrum. The record of each photograph was analyzed as the remaining light reflection after blocking the designated wavelengths of light. At the same time, this is the prepress

that can prove the authenticity of the document in the case of forgery. Textual information can be found on the document: "SECURITIES CONFIRMATION". Hidden information is programmed as vector graphics enveloped in the dense lines of different sets of twin colorants. Since the design is applied in two light spectra, it has two images which are observed separately with ZRGB cameras. Images V (visible) and Z

(NIR spectrum) present an example of security graphics with mutual concealment if they are printed next to each other and partially nested in each other. Although the original design was created as vector graphics, the prepress is a high-resolution pixel graphic.

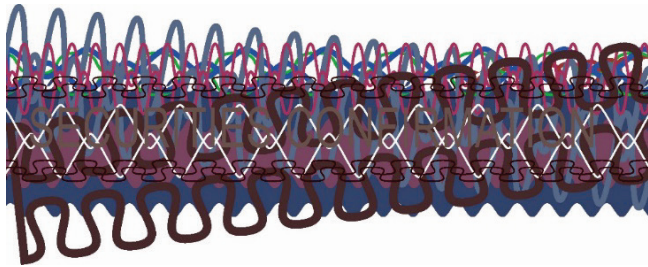
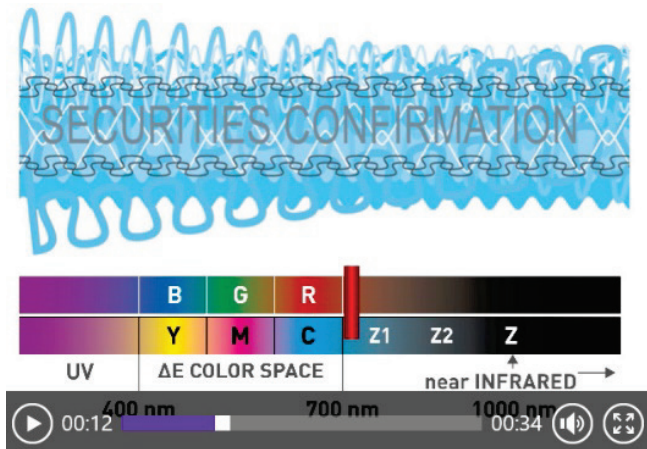


Figure 1 Security graphics in the visible spectrum

The transition of the color twins and change in the light absorption from the V to Z state in dual presentation are shown as a video animation. Fig. 2 shows security graphics in the blockade at 700 nm. Some colorants already disappear in the visible spectrum, while others are visible up to 750 nm. After 800 nm, only the information on the absorption of the near-infrared spectrum remains.

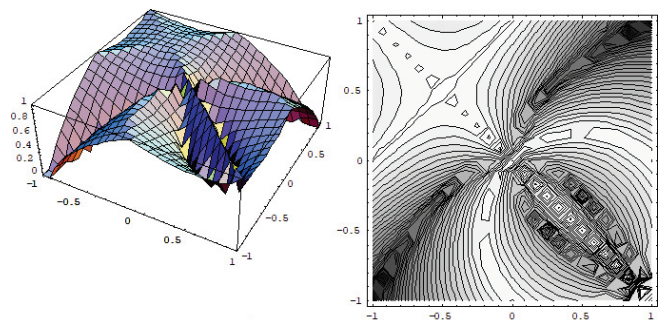


<http://www.jana.ziljak.hr/securityTG.mp4>

Figure 2 Twin dyes in blockages from the visible to the NIR spectrum up to 1000 nm.

3 SCREENING FOR SECURITY PRINT WITH THE MUTATING RASTER FORM R87

Individualized rasterization is used in design as an original solution. This is an introduction into security graphics with an emphasis on the stochastic choice of line screen coverage [10]. In this paper, the mutating screen form R87, which deforms and changes its appearance in print depending on the amount of colorant applied, is presented. After pixelization, raster color interpretation is introduced. With them, the recognition of color-coded twins that are nested is reduced.



$$1 - \text{Abs} \left[-\sqrt{\text{Abs}[x^2 + y]} + \sqrt{\text{Abs}[x - y^3]} \right]$$

Figure 3 The model of the screen form R87 in 3D/2D with the corresponding algorithm

Table 2 Algorithm (PostScript – GSview 5.0) of the activated three-screen shape:

```
15 600 translate 40 80 scale
/rounddot {dup mul exch dup mul add 1 exch sub} bind def
/propeller {dup 2 index 3 exp sub abs sqrt 3 1 roll 3 exp add abs sqrt exch
sub abs 1 exch sub} bind def
/R78 {dup 2 index 3 exp sub abs sqrt 3 1 roll 2 exp add abs sqrt exch sub
abs 1 exch sub} bind def
/lin 5 def /kut 0 def
lin kut {rounddot} bind setscreen 0 1.1 neg translate
10 1 4 [1 0 0 1 neg 0 0] {<3456789abc>} image
lin kut {propeller} bind setscreen 0 1.1 neg translate
10 1 4 [1 0 0 1 neg 0 0] {<3456789abc>} image
lin kut {R78} bind setscreen 0 1.1 neg translate
10 1 4 [1 0 0 1 neg 0 0] {<3456789abc>} image
lin kut {R78} bind setscreen 0 1.1 neg translate
10 1 4 [1 0 0 1 neg 0 0] {<8394c26b5a>} image
Showpage
```

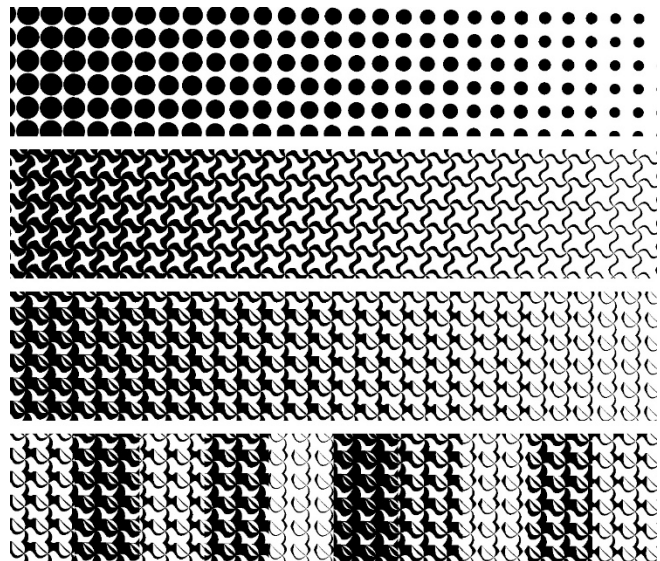


Figure 4 Macro-view (5 lines per inch) of the continuous screen/raster form: round dot, propeller, R87, and 4 bit hexa schedule <8394c26b5a> R87

The first raster shape is a standard round dot applied in conventional printing. For security printing, we suggest new raster forms: "propeller" and R87. With the passage of time and the ageing of documents, the colorant values change. The IRD technique insists on minimizing the difference between the twins of colors in the visible spectrum. The ΔE numerical values are less than three, which is recommended for the

mutual concealment of two graphics. The experimental work in the development of twin colorants has confirmed that mixing the V and Z twins of colorants in print depends on the geometrical shape of the raster element. Better solutions for hiding the V and Z images are achieved with mutating rasters from the "horn-shaped" series. The Z graphics printed next to the V graphics (with the same color) are less discernible if their contact surfaces are jagged. This was the motivation to create and apply the new raster form R87, resulting in an individual solution which raises the quality of print in security graphics.

4 NIR SPECTROSCOPY OF TWIN COLORANTS FOR THE OFFSET PRINTING TECHNIQUE

The equality of the twins of colorants in the visible spectrum is expressed through the ΔE value, for which several analytical solutions have been proposed [11]. The ΔZ value, which measures the difference of light absorption in the near-infrared spectrum [12], has been introduced into the IRD printing technique. There is no analytical solution for this value. The practices of photographing with ZRGB cameras and determining the difference of grayness on photographs with conventional programs for image processing are introduced. In this article, it is shown that the spectrum of light absorption for the twin colorants in the visible and near-infrared spectra is from 400 to 1000 nm. Spectrograms are given for colorants from the examples of security print (Fig. 2), with an emphasis on the Z value of the wavelength area at 900 nm. Fig. 6 shows spectrograms of the twins of the gray and green color tone. Spectroscopy was performed from the print samples.

The dashed lines (V colorants) drop to zero after 800 nm in the NIR region. At a measurement above 900 nm, the difference between the V and Z colorant twins is determined. Dyes with the K content greater than 40% retain the ΔZ range that the Z camera will interpret as a dark area.

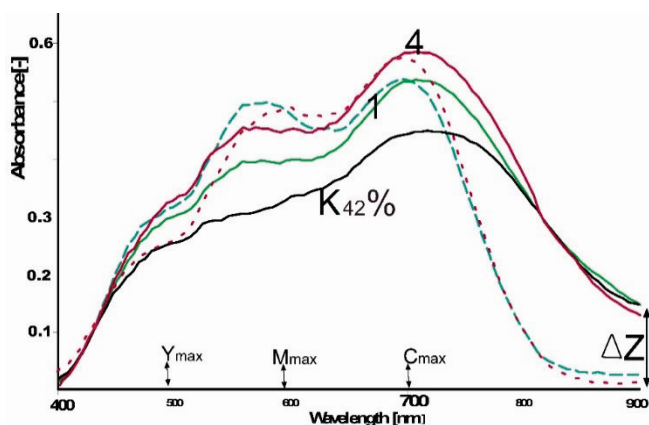


Figure 5 Spectrograms of the light absorption of colorants 1, 4 (Tab. 1)

The light absorption maxima are: yellow at 490 nm, magenta at 590 nm and cyan at 700 nm [12]. The advantage of the spectral, graphical representation is explained by the twins' dyes No. 1. There are major discrepancies in the colors of the twins No. 1 in the magenta area. This is information on

the necessary repair of the dye mix No. 1. The difference in the absorption value in the magenta dominance (from 540 nm up to 650 nm – 69% / 47%) is too large. The difference in the magenta coverage in twins with 42% K is too small. In the next iteration of the trial print, either the magenta V dyes (solid line) will increase or the magenta content of the Z twin (dash line) will decrease. The line No. 2 (Fig. 6) is a thin dark line with a high proportion of carbon black dyes (65%). That line will remain in the final Z image (video, Fig. 2)

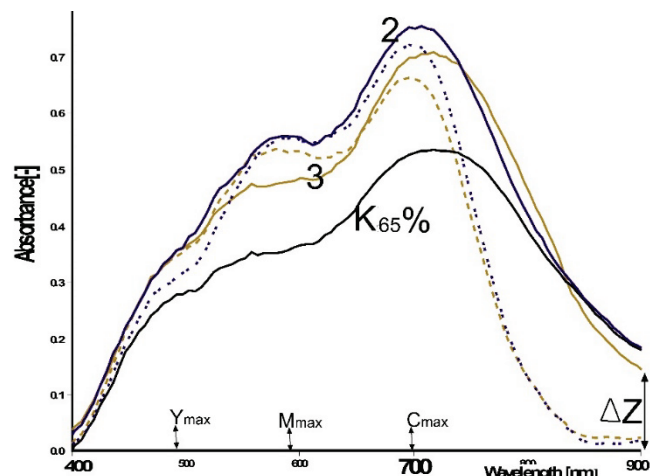


Figure 6 Spectrograms of the light absorption of the colorants 2, 3 (Tab. 1)

The parameters for the ratio of the C, M, Y, K process components were determined in six iterations of print. The graphic representation of light absorption is the ideal method for proposing a new direction in adjusting the ratios of every CMYK component. Since process colorants are in their own domains, spectrograms of individual components as the final optimal solution are given in the article [12].

5 DISCUSSION AND CONCLUSION

In our work with security graphics, the colorant carbon black for the Z twin is given beforehand. Security graphics in our research use colorants with a strictly defined order: the NIR colorant has coverage of 40% and is determined in advance. All other parameters in the V and Z colorants are subordinate to this requirement for security print. The amounts of C, M, Y, K colorants in V and Z are adjusted in order to minimize ΔE . Such design of the relationship between the V and Z graphic is the beginning of developing contrast in the instrumental observation of the visible and hidden image. A larger amount of the carbon black colorant reduces the scope of the coloration of security graphics. We propose further experimentation with smaller values of the coverage of the component K in Z graphics.

Z graphic comes at the end of the graphic postpress for the four-color print, because vector graphics programs ignore it in the second, invisible order. Joining the lines, the overlay order is carefully worked out because the final graphic is interpreted as pixel graphics. Discarding information (second line "below") is a standard process in the digital transition from vector to pixel graphics. Let us emphasize: this is not

about the transparency of dyes. Only, if press were released separately for each layer, printing with spot dyes would be possible. This is possible in offset, but the printing is significantly more expensive. A test print on a digital printer is not suitable for "print in layers".

Security graphics recognizes different process colorants based on their absorption of near-infrared radiation. Three colorants for standard offset printing – cyan, magenta and yellow – do not absorb light outside the visible spectrum. We plan to use them as components in graphics which are intended only for the visible spectrum without showing up in the near-infrared spectrum, but with set differences in the V and Z graphics. This proposition is the beginning of creating security graphics with two different levels of content, which are instrumentally separated by using ZRGB cameras.

IRD security graphics is sensitive to the quality of twin colorants and their sameness in the visible spectrum. We propose the introduction of raster elements with pointed forms. The mixing of colorants and the contact between the V and Z graphics is planned at the level of an irregular raster form. Such vibrating graphics have two advantages. Firstly, they are almost impossible to detect even with high-resolution scanners. Scanners that are currently in use reduce all images to RGB records, which are then interpreted using the GCR (CMYK) procedure, destroying the principle of the IRD print in the process. Furthermore, the two graphics of V and Z are in contact; they have a different content of the same color at the same place. For the naked eye, the new raster elements (R87) reduce the gap, i.e. the possibility to recognize the boundary between the graphics V and Z.

The planning of the quantities of individual components of process colorants is observed through the ΔE and ΔZ values. The amount of each CMYK component in twin colorants is determined through the spectrography of the mixture and the spectrography of each CMYK component separately. A final note: the new printing system of security graphics is based on the use only with the same colorants, which are used for conventional print.

7 REFERENCES

- [1] Friščić, M., Agić, A., & Žiljak Stanimirović, I. (2017). Visual and infrared graphic applied through dedicated halftoning for transparent polypropylene packaging. *Tehnički vjesnik*, 24(1), 225-230. <https://doi.org/10.17559/TV-20151231105549>
- [2] Anayath, Rajendradrakumar & Žiljak, V. (2011, Feb.). Invisible pics hit newspaper. *RIND Survey*, Chennai, India, Rs.40.00, 32(2), 4-6.
- [3] Žiljak, V., Akalovic, J., & Žiljak-Vujić, J. (2011). Upravljanje bojilima na koži u vizualnom i infracrvenom spektru. *Tekstil : časopis za tekstilnu tehnologiju i konfekciju*, (0492-5882), 60(8), 355-363.
- [4] Žiljak, V., Pap, K., & Žiljak, I. (2009). CMYKIR security graphics separation in the infrared area. *Infrared Physics and Technology*, 52(2-3), 62-69. <https://doi.org/10.1016/j.infrared.2009.01.001>
- [5] Žiljak Vujić, J., Rajković, I., & Žiljak Stanimirović, I. (2014). Simultano video snimanje u vizualnom i infracrvenom spektru proširene V/Z stvarnosti. *Polytechnic & Design*, 2(1), 73-78.
- [6] Rajković, I. & Žiljak, V. (2018). Parallel Motion Images in Visual and Near Infrared Spectrum. *Tehnički vjesnik*, 25(4), 1004-1008. <https://doi.org/10.17559/TV-20160610125726>
- [7] Žiljak, J., Jurečić, D., & Žiljak, V. (2018). Packaging Design with Hidden near Infrared Colour Separation. *Tehnički vjesnik*, 25(3), 211-215. <https://doi.org/10.17559/TV-20170705114921>
- [8] Žiljak, V., Pap, K., & Žiljak-Stanimirović, I. (2011). Development of a Prototype for ZRGB Infraredesign Device. *Tehnički vjesnik*, 18(2), 153-159.
- [9] Projectina Docucenter 4500, PAG B50 <http://forensictechnology.com/projectina/> <http://www.projectina.ch/cre8.upload/pdfs/UCM%20LED%20PIA-7000%202011-e.pdf>
- [10] Žiljak Stanimirović, I., Anayath, Rajendra, & Bogovic, T. (2010). The Infraredesign with Individualised Screening. *Proceedins of 11th International Design Conference, Design Graphics with security elements*, Vol. 4, The Design Society, Dubrovnik, Croatia, 1863-1868.
- [11] Glogar, M. I. & Parac-Osterman, Đ. (2015). Achromatic Hues Matching in Graphic Printing. *Acta graphica*, 26(1-2), 36-45. <https://hrcak.srce.hr/145334>
- [12] Žiljak Gršić, J. (2017). Near infrared spectroscopy in print technology. *Polytechnic & Design*, 5(1), 32-36. <https://doi.org/10.19279/TVZ.PD.2017-5-1-05-en>

Authors' contacts:

Vilko ŽILJAK

(Corresponding author)
Faculty of Graphic Arts, University of Zagreb,
Getaldićeva 2, 10000 Zagreb, Croatia
vilko@ziljak.hr

Jana ŽILJAK GRŠIĆ

Zagreb University of Applied Sciences,
Vrbik 8, 10000 Zagreb, Croatia
jana@ziljak.hr

Denis JUREČIĆ

Faculty of Graphic Arts, University of Zagreb,
Getaldićeva 2, 10000 Zagreb, Croatia
denis.jurecic@grf.hr

Tonči JELIČIĆ

Hydrographic Institute of the Republic of Croatia,
Zrinsko-Frankopanska 161, 21000 Split, Croatia
tonci.jelicic@hhi.hr

IMPLEMENTATION OF INTELLIGENT MODEL FOR PNEUMONIA DETECTION

Željko KNOK, Klaudio PAP, Marko HRNČIĆ

Abstract: The advancement of technology in the field of artificial intelligence and neural networks allows us to improve speed and efficiency in the diagnosis of various types of problems. In the last few years, the rise in the field of convolutional neural networks has been particularly noticeable, showing promising results in problems related to image processing and computer vision. Given that humans have limited ability to detect patterns in individual images, accurate diagnosis can be a problem for even medical professionals. In order to minimize the number of errors and unintended consequences, computer programs based on neural networks and deep learning principles are increasingly used as assistant tools in medicine. The aim of this study was to develop a model of an intelligent system that receives x-ray image of the lungs as an input parameter and, based on the processed image, returns the possibility of pneumonia as an output. The implementation of this functionality was implemented through transfer learning methodology based on already defined convolution neural network architectures.

Keywords: computer vision; machine learning; neural networks; pneumonia

1 INTRODUCTION

Nowadays, machine learning and artificial intelligence methods are culminating in every aspect. These methods are increasingly being used in business systems, enterprises of various types and in science, in order to improve efficiency of the facilities they serve. While numerous works focus on a conscious form of artificial intelligence that would somehow replace humans, computers with learning ability have been around for some time.

Object recognition and deep analysis are just some of the characteristics of machine learning and neural networks. Although the concept of neural networks was developed back in the 20th century, training process represented hard job for computers from that time. With the development of hardware, GPU (Graphical Processing Unit) and CPU (Central Processing Unit), and high availability of data, neural networks are experiencing significant upswing. Sometimes the diagnosis of the disease was almost exclusively dependent on the experience of the doctor and his assessment. Today, more and more diseases can be diagnosed with the help of some kind of computer or machine.

This paper is based on the diagnosis of pneumonia according to the obtained x-ray image as an input parameter. With the advancement of artificial intelligence, it can be very useful in the diagnosis of x-ray images and play an important role in the detection of disease, serve as a computer assistant to radiologists, and increase their efficiency in diagnosis.

Particularly noteworthy is the application of AI system for the diagnosis of pediatric pneumonia using chest X-ray images. This tool may eventually aid in expediting the diagnosis and referral of these treatable conditions, thus facilitating earlier treatment, resulting in improved clinical outcomes. [1]

The aim of this project is to implement software for the automatic detection of pneumonia and the use of computer algorithms in the field of machine learning to automate the process of obtaining the most accurate diagnosis, which could reduce the possibility of errors and misdiagnosis that

can lead to unwanted consequences. Project uses several different technologies combined. The implementation of the convolution neural network model was done in Python using Anaconda as programming environment. [2]

2 MACHINE LEARNING

Machine learning is a branch of artificial intelligence that focuses on designing algorithms that improve their own performance based on input. Machine learning is one of the most active and represented areas of computing and is the base of data science.

It enables computers to learn in a manner similar to humans, that is, to gather knowledge based on past experience and analysis. Instead of constantly updating the program code, the computer eventually becomes independently capable of improving the performance of its algorithms. Data processing using machine learning methods often results in a model capable of performing some kind of prediction on later test data. [3]

2.1 Deep Learning

In this paper, the focus is on deep learning, which can be characterized as a branch of machine learning inspired by the structure of the human brain. This principle was developed by modeling the first neural networks back in the 1940s, with the term first mentioned by Rina Dechter in 1986 and defined as a field of machine learning based on the presentation of data to complex representations at a high level of abstraction to which comes by learned nonlinear transformations.

Deep learning methods are most commonly used in challenging areas where the dimensionality and complexity of the data is extremely high. Deep learning models mainly carry out learning process on a large set of data, which falls within the scope of supervised learning, and form the backbone of today's principle of autonomous driving, disease detection and pattern recognition, with result that were impossible before. [4]

2.2 Biological Neural Networks

The human brain consists of densely interconnected nerve cells that serve to process data, also called neurons. Within the human brain, there are approximately 10 billion neurons and approximately 60 billion connections between them.

A neuron as a unit represents a very simple structure, but a large number of neurons represent tremendous processing power. One neuron consists of a soma, representing the body of a cell, fibers called dendrites, and one longer dendrite, or axon. Dendrites are located in a network around the cell of that same neuron, while the axon extends to the dendrites and cells of other neurons. Signals or information are transmitted from one neuron to another by complex electro-chemical reactions. Chemicals are released from synapses and cause a change in electrical potential in neuronal cells. As the electric potential threshold is reached, the electrical impulse sends the action potential over the axon. The impulse extends until it reaches a synapse and increases or decreases its potential.

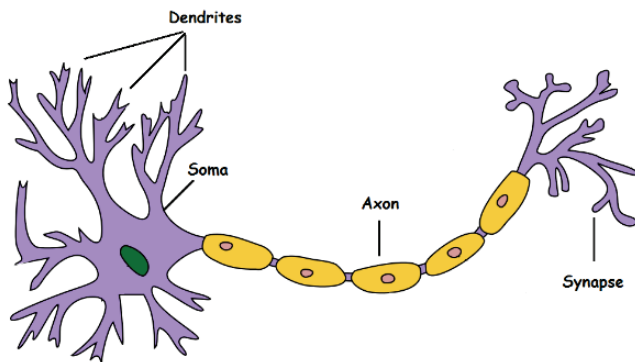


Figure 1 Structure of a biological neuron

2.3 Artificial Neural Networks

The artificial neural network, inspired by the work of the human brain, contains a number of connected processors, neurons, which have the same role as biological neurons within the brain. They are connected by links whose signals can pass from one neuron to another, thus transmitting important information. Each neuron receives a certain number of input signals in the x_i tag. Each connection has its own numerical value, namely the weight of the w_i , which is the basis for long-term memory in artificial neural networks. The x_i and w_i values are summed by the transfer or activation function, and the result is sent as output Y to another neuron.

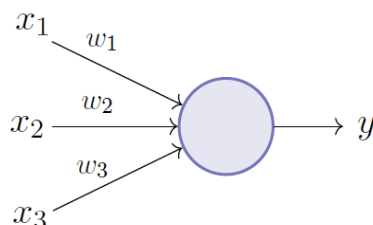


Figure 2 Structure of artificial neuron

In artificial neural networks, neurons are arranged in layers that are interconnected by the links through which the signals pass, that is, information that can be of great importance for the end result and performance of the neural network. Connections between neurons are triggered if the condition is met. Condition is defined with activation function that will be elaborated later.

The layer that receives information from the environment is called the input layer. It is associated with the hidden layers in which information is processed, that is, the network learns. The last layer that produces outputs is called the output layer. The learning process is performed by changing the value of the weights or connections between the neurons, comparing the required and obtained sizes on the output layer and calculating the error.

Based on the error value obtained, it is attempted to be reduced by each subsequent step, and this is done by weight correction based on a defined learning rule related to data acquisition.

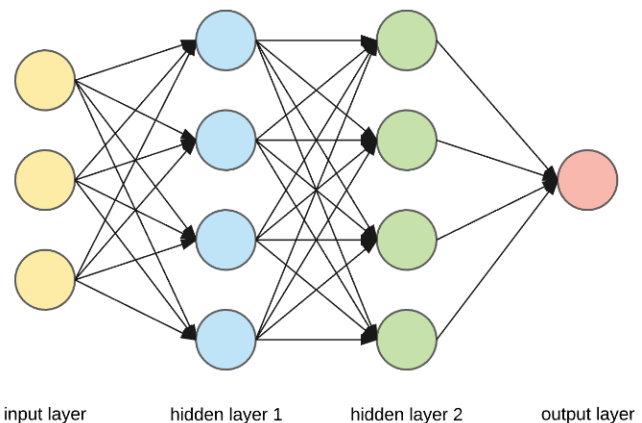


Figure 3 Layers of artificial neural network

2.4 General Learning Rule

Each artificial neural network is based on a general learning rule that involves collecting relevant data, which is then divided into training and validation data.

After the data collection is completed, it is necessary to determine the architecture of the neural network, which involves determining the number of layers in the network and the number of neurons in each layer, and then selecting the type of connection between the neurons together with the learning rule that is the basis for defining criteria that determines the architecture of the neural network. The next step is learning, which is the basis of artificial neural networks. Learning involves initializing weights, training a model based on a training dataset, and checking the amount of error that weights are corrected after each iteration, which refers to going through all the training samples, called the epoch.

Learning lasts until the desired number of epochs is reached or a wanted error is met. In the initial stages of learning, the neural network will adapt to the general trends present in the data set, so the error in both the training set and

the validation set will fall through time.

During the longer learning process, there is a possibility that the neural network can begin to adapt to the specific data and noise of the learning data, thereby losing its generalization property. The error on the training set will drop while on the validation set it will start to grow. The moment the validation set error starts to grow, the learning process must be interrupted so that the model does not become overfitted.

With the completion of the learning process, it is necessary to test the operation of the network using previously obtained validation data. The difference between the learning and the testing phase is that the neural network no longer learns in the testing phase and the weight values are fixed. The evaluation of the network is obtained by calculating the error and comparing it with the desired error value. If the error is greater than allowed, additional training data may need to be collected or the number of epochs increased for better results, since in this case the network is unsuitable for use.

2.5 How Neural Networks Work

In working with neural networks, there are two basic algorithms used today. These are the feedforward algorithm and the so-called backpropagation algorithm.

Working with neural networks allows us to define complex, nonlinear hypotheses consisting of one or more neurons that can be arranged in one or more layers that build the neural network architecture. This method of mapping the input vector to the network outputs is called forward propagation.

Backpropagation algorithm is one of the main reasons that made artificial neural networks known and usable for solving different types of problems. With this algorithm, artificial neural networks have been given a new capability, supporting multiple layers, and the backpropagation algorithm has proven to be the most common and effective method in deep neural networks. The algorithm is used in conjunction with optimization methods such as gradient descent. The principle of network operation is based on the transfer of input values from the input to the output layer, determination of the error and propagation of the error back through the neural network from the output to the input in order to train the network as best as possible and to reduce the existing error and thus improve the final results.

2.6 Types of Activation Functions

There are many types of activation functions that determine whether and how neurons are activated within a network. Activation functions can be divided into linear and nonlinear, and in practice, only a few that have proven useful are used. Linear functions are used in regression problems when unlimited output of any kind is required in the output layers, while nonlinear activation functions are more suitable for working with classification problems where the outputs are limited to small quantities.

The most popular functions to use within classification

problems are jump functions and sigmoidal functions.

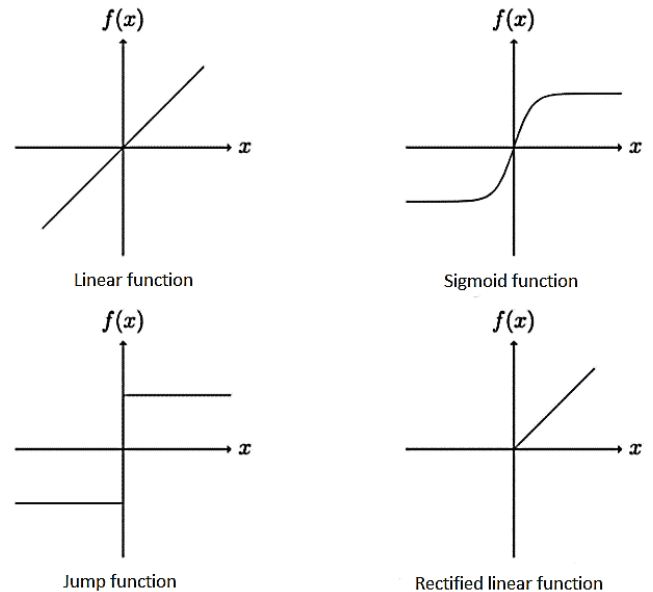


Figure 4 Types of activation functions

2.7 Convolutional Neural Networks

In the neural network structures mentioned, each neuron output was a scalar value. Convolutional neural networks are a special type of artificial neural networks for processing unstructured data such as images, text, sound and speech in which extensions occur in the form of convolution layers. The outputs from the convolution layers are two-dimensional and are called feature maps. The input to convolutional neural networks is two-dimensional (image data), while kernels are used instead of weight values. In addition to the convolution layers mentioned, these types of neural networks have specific pooling layers and fully connected layers.

Convolutional neural networks usually start with one or more convolution layers, followed by a pooling layer, then the convolution layer again, and the process is repeated several times.

The convolution layer is a fundamental part of any convolution neural network. Each convolution layer consists of filters containing weight values that need to be learned in order for the network to return better results.

In the initial phase, the input convolves (multiplies) with the filter and produces a scalar product over the entire width and length of the image. The result of the convolution is a two-dimensional activation map that shows the filter response at each position in the array. In this process, the network will learn the weights within the filter to activate the filter where it recognizes certain image properties, edges, shapes, and the like. In order to define the exact size of the output of the convolutional layer, it is also necessary to define the stride of the filter according to the input data. The stride determines how far the filter will move in height and width during the convolution process.

The output width is indicated by W_y and the input width by W_x . The filter stride is denoted by the letter S , while F

indicates the square filter size. The output width after the convolution can be defined as:

$$W_y = \frac{W_x - F}{S} + 1, \quad (1)$$

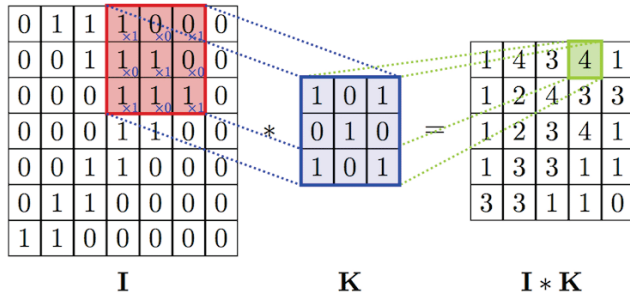


Figure 5 Convolution process

The pooling layer contains a filter that reduces the dimensions of an image. In a convolutional neural network, pooling layer is most commonly used after several convolution layers in order to reduce the resolution of maps generated by the convolution layer. The pooling layer filter is different from the convolution layer filter because it does not contain weight values. The specified filter is used to select values in the filter default dimensions.

Of the several types of pooling, the most commonly used are average pooling and max pooling. Average pooling replaces the arithmetic mean of clustered values, while the max pooling simply selects the maximum value. The benefit of max pooling is to store the stronger and more prominent pixels in the image, which are more important for getting the end result, while the irrelevant pixels are eliminated.

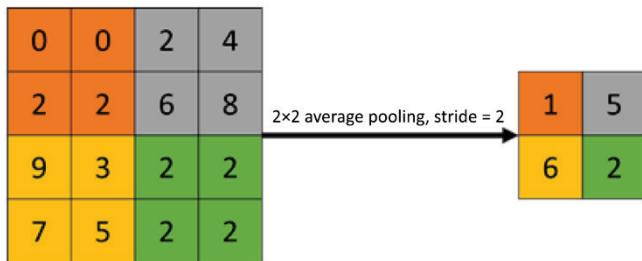


Figure 6 Average pooling

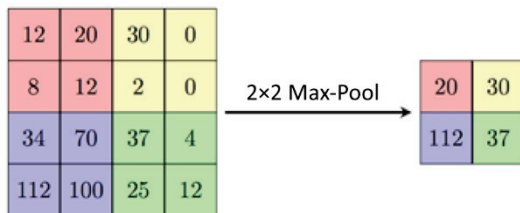


Figure 7 Max pooling

In the pooling layer, the dimensions of the output maps are calculated as:

$$D_n = \frac{D_s - F}{S} + 1, \quad (2)$$

where D_s is the old dimension, F is the filter width, and S is a jump between 2 value selections.

The fully connected layer is in most cases used in the final layers of the network. The reason for using fully connected layers is to reduce the dimensions of the image by passing it through the neural network, since complete connectivity is defined as the square number of connections between layers.

For example, for image data with dimensions $200 \times 200 \times 3$, the input layer would have 120,000 input values. If we fully linked this to a hidden layer consisting of 1000 neurons, we would have 120 million weight values to learn, which requires big computing power. This is why fully connected layers are used in the later stages of the neural network.

2.8 Avoiding Local Minimums

During the learning process of the neural network, the goal is to find the location of the global minimum of error, which means that the model is at the best possible level at a given moment, and the learning process can stop. In this process, the so-called local minimums can fool the network in a way that the network thinks it is within the global minimum. Avoiding local minimums can be achieved by using various methods.

Known methods for avoiding local minimums:

Random Transformations – Random transformations serve to augment an existing training dataset that is accomplished by operations such as translation, rotation, and scaling. This increases the number of data without the need to collect additional samples. An increased amount of learning data makes the network less likely to get stuck in local minimum. Random transformations can be performed during each iteration in the learning process or by pre-processing data before training begins.

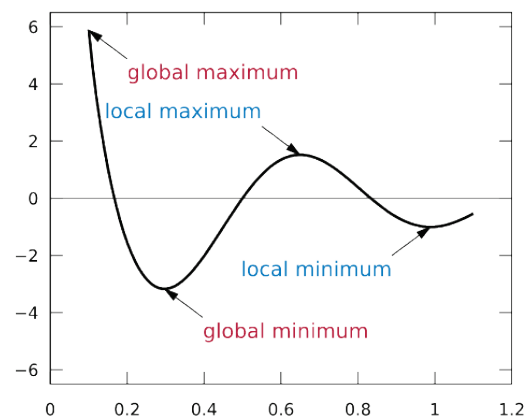


Figure 8 Local and global minimum comparison

Dropout – During every iteration of the learning process, some neurons are "accidentally" switched off, thus creating the appearance of a large number of architectures, although in reality they only work with one. In the case of multiple architectures, it is unlikely that they will be stuck at the same local minimum.

2.9 Transfer Learning

Conventional deep learning algorithms are traditionally based on isolated tasks, while a single neural network serves a particular type of classification. Transfer learning is a new methodology that seeks to change this and circumvent the paradigm of isolated learning by developing knowledge transfer methods that can use models learned on one type of classification for multiple different tasks.

In this way, a model initially created for one type of problem can later be used as a starting point for solving a new type of classification, and thus give better results than initializing a new neural network from the beginning.

An analogy can be made with learning to drive in real life. In this analogy, learning how to drive a motorcycle can be greatly assisted by the knowledge of driving a car. Transfer learning works in a similar way.

Transfer learning is nowadays a very popular approach on which the practical part of this paper is based. Choosing pre-trained models and already defined neural network architectures can be of great use in solving complex problems such as detecting pneumonia on x-ray images.

After the architecture of the existing neural network with defined layers has been loaded, it is necessary to delete the last, output layer from the specified network, and replace it with a new output layer related to the problem for which we will use the network.

The next step is to conduct training on the complete, already defined, architecture and all layers of the network on the learning dataset. By using this way of learning, the neural network will be able to apply the classification principles learned in the previous tasks to a new type of problem, and in that way the results will be better without the need to define layers and create a new neural network from the beginning.

Transfer learning is good to use in cases where we do not have a large dataset to learn. In the case where we have approximately 1000 images to perform learning process, by merging that 1000 data with trained networks that have been trained with millions of data, we can gather many learned principles for sorting and classification, and in that way improve model efficiency and reduce training time.

VGG16 is a convolution neural network architecture proposed by K. Simonyan and A. Zisserman of Oxford University. The model achieves 92.7% accuracy on the ImageNet image dataset, which consists of over 14 million image data divided into 1000 classes.

The VGG16 network architecture consists of 13 convolution layers in which the number of filters increases from 64 to 512; 5 compression layers with highest value compression; and, 3 fully connected layers that are used to avoid local minimums using the dropout technique described earlier.

Big progress in comparison with previous neural network architectures has been made after resizing the convolution filter to 3×3 .

The filter of the pooling layers is 2×2 in size with a stride of 2, while all the hidden layers of the network use the rectification linear activation function described in the previous chapters. All convolutional layers use stride and

padding of value 1, and the total number of parameters of the specified network architecture is 138 million. [5]

3 MODEL IMPLEMENTATION

This chapter describes the implementation of the neural network model. The process of collecting image data for learning, division of the set into a learning and validation set, visual analysis and comparison of data for learning and with data for validation, comparison of the number of positive and negative images, and preprocessing of the collected data for entering the learning process are presented.

The procedure of retrieving the previously described architecture of the VGG16 network and changing the output layer in accordance with the collected data were developed, the learning curves using Tensorboard technology and the implementation of the training or learning process were presented. After learning was completed, the best-performing model was saved and an evaluation of that model was performed to graphically present prediction accuracy on validation data not used in the learning process.

3.1 Data Collection

The first step in implementing a model for detecting pneumonia on x-ray images is to collect imaging data that will consist of sets for training and validating the model. Dataset available for download on the Kaggle site was used. Kaggle offers users a large number of different datasets that can be used for various research purposes.

The dataset selected consists of a total of 5863 lung x-ray images divided into two categories. Existing categories are x-rays with a positive diagnosis of pneumonia and images of normal, healthy lungs with no indication of disease. The set contains 1583 images of healthy lungs and 4273 images positive for pneumonia. The image data format is .jpeg, while the image dimensions are different and vary from image to image. In the later steps, it is necessary to change the image dimensions that will be supported as input for the VGG16 network architecture. [6]

3.2 Data Preprocessing

After successful process of the data collection for training process, before learning, the same data should be processed in such a way that they are suitable for entering the network.

Image file paths were originally loaded using a function from the Scikit-learn library that loads the paths of all files in a given directory as parameters of those directories within the main directory as categories of individual files, or their paths.

Subsequently, categories or labels are defined from the obtained data set, depending on whether the data is in the subdirectory of images with pneumonia or images without disease. Saving image categories was done using functionality from the Numpy library.

By completing the label definition for all retrieved image paths, the total image dataset should be divided into a learning set and a set for validation. In this case, 90% of the

total data set was determined for learning purposes, with the remaining 10% for validation.

After splitting paths and labels into training and validation data, functions were implemented to load saved paths and convert them to image data and resize images to 256×256, which will be required for later input to the initial layer of the neural network.

3.3 Data Visualization

This chapter elaborates a section based on the visualization of preprocessed data for an easier idea of the dataset collected, the ratios between learning and validation data, and the like. Graphics and diagrams to show the ratio between the data were implemented using the Matplotlib library.

A visualization feature that has been implemented offers a graphical representation of the data. This function performs visual processing of the data ratio, namely, the ratio between the learning and the validation sets and the number of images positively diagnosed for pneumonia in relation to x-ray images of healthy lungs.

A visual representation of the ratio of the number of learning data sets to the validation set:

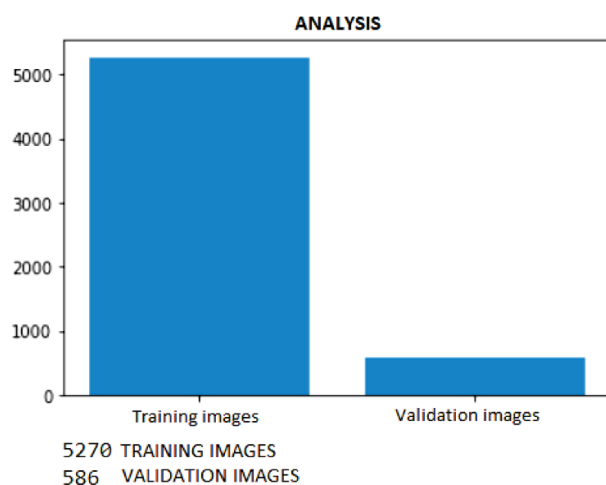


Figure 9 Comparison of validation and training data

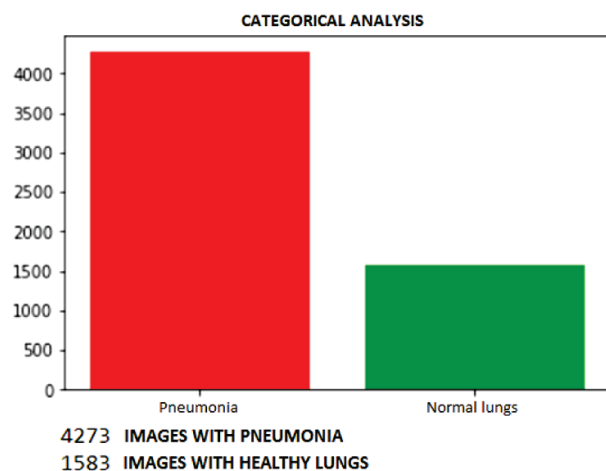


Figure 10 Comparison of positive and negative samples

3.4 Model Architecture

Defining a convolutional neural network architecture is the most important aspect on which the overall success of a project depends. This part of the paper was done using the transfer learning methodology described earlier in the paper, which uses defined architectures available for use in order to have a better model performance. The VGG16 architecture was used in this paper. In this way, the neural network can take advantage of all previously learned principles such as recognizing edges, angles, discoloration, etc., thus shortening learning time and improving model efficiency.

The VGG16 architecture was loaded using a function from the Keras library. Another layer of average pooling layer was added to the defined VGG16 architecture. The last layer of the neural network was changed, which defines the number of classes of possible outputs, in our case 2, that is, pneumonia and normal state of lungs. A softmax activation function has been added to the last layer, which will change the output values depending on the ratio in values ranging from 0 to 1 in order to easily obtain a percentage of the prediction value.

The initial VGG16 architecture and the changes on the last layer are integrated into a single architecture, after which the complete architecture is stored in a variable that will execute the learning process. The functions for calculating the error, the type of optimizer with the learning rate and the values to be monitored during the iterations of the learning process are defined, in our case the prediction accuracy and the error value is included by default. The use of Tensorboard enables a later overview of the learning flow and changes of defined parameters using Tensorboard technology.

After completing all of the above procedures, the model architecture has been successfully implemented and is ready to perform a model training process that may take some time depending on the computer configuration and the amount of learning data.

3.5 Training Process

The process of learning a defined model is the part where the so-called neural network magic takes place. The model will in an iterative way, using the backpropagation algorithm, learn which weight values are best and most effective for producing a model that can detect pneumonia. Once the convolutional neural network architecture has been implemented, the path of the model in which the weighted learning values will be stored is determined. An option is selected that allows only the best model to be saved, to save memory, and to avoid storing weight values at each iteration. The fit function in the domain of Keras technology is called to perform learning process. Images and labels from the learning set and the validation set are sent as parameters, epochs, checkpoints for storing weights, and Tensorboard for later graphical representation of learning curves after learning are determined.

3.6 Model Results

A learning curve that shows the evolution of model accuracy by epochs and the decrease in the amount of error is presented using Tensorboard technology. Graphs were drawn to visualize the accuracy of the model on the learning set across epochs, the error drop on the learning set, and the accuracy percentages and the error drop on the validation set. The tensorboard can be open by typing the `tensorboard --logdir = logs /` command in the Anaconda command line. The word `logs /` indicates the directory where graphs are stored. The localhost: 6006 address is then entered in the browser and the Tensorboard window in the browser opens. Graphs related to the results obtained from the learning set:

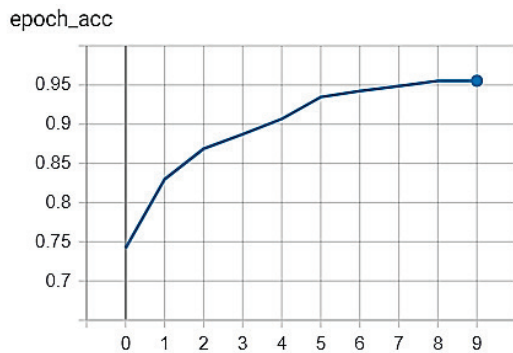


Figure 11 Accuracy curve on learning set

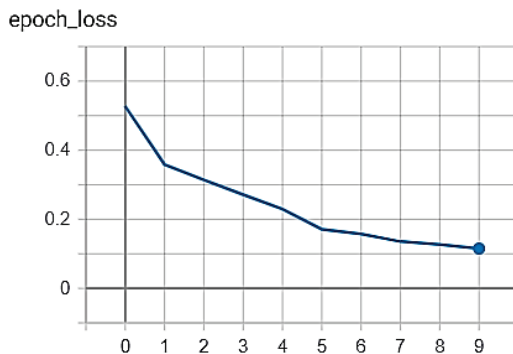


Figure 12 Error curve on learning set

The curves produced from the learning dataset show us an increase in accuracy across all 10 learning epochs, while the error decreases, which is the goal of the model presented. A much more important aspect of the model is the results based on a validation set that has not been used in the learning process to see the model's response to unprecedented data. Graphs from validation set of x-ray image data are shown in the Figs. 13 and 14.

The results on the validation dataset show an overall accuracy of 94%, which is an indicator of the high performance of the pneumonia prediction model. [7]

After the epoch that was one before the last one, it is noticeable that the amount of error at the validation set starts to increase, while at the training set it continues to decrease, which emphasizes the possibility of overfitting. It is necessary to stop the learning process after the 10th epoch, which proves the optimal value of the number of epochs for

this type of problem.

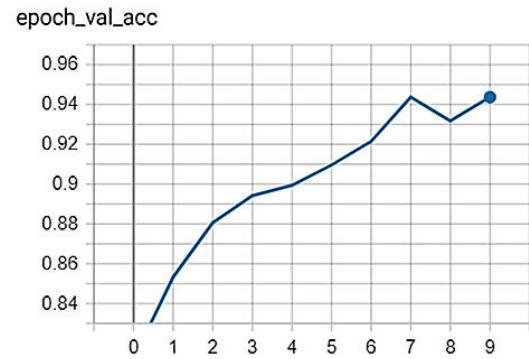


Figure 13 Accuracy curve on validation set

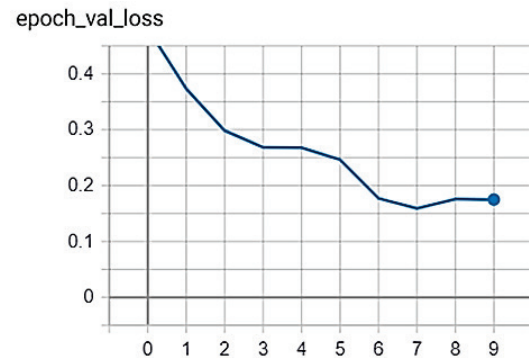


Figure 14 Error curve on validation set

After visualizing the learning process, a confusion matrix was used to display the number of correctly and incorrectly classified images on a validation dataset divided into two classes. The machine learning confusion matrix serves as an instrument for conducting the evaluation of the learned model in the domain of classification problems.

After graphical representation of the confusion matrix, using the Keras evaluate function, the overall accuracy of the model on the validation dataset is 94%. Accuracy is determined on the validation dataset because in the case of evaluation of the learning set, overfitting may occur, and in that way the true accuracy of the model cannot be determined.

The 94% accuracy rate indicates that the model is very well trained on a relatively small dataset of 5000 images. A very important factor in this is transfer learning, in which the architecture of the VGG16 network uses some of the previously acquired knowledge and achieves high accuracy in categorizing pneumonia.

The Fig. 15 shows a visualized confusion matrix created as a product of the steps described above.

By analysing the confusion matrix shown, the high accuracy of the model can be noticed. Of the 429 test RTG images with pneumonia, 415 images showed pneumonia, while only 14 images were misclassified.

In the test dataset, there are significantly fewer images without pneumonia, and out of a total of 157 images, 138 were correctly classified, and for 19 images the model gave incorrect results.

The matrix shows a diagonal shape in darker colours, which is a good sign since the confusion matrix with the strength of its diagonal shows the superiority of the learned model. The total accuracy of the model is below the matrix when all the images are combined together, and it shows the 94% accuracy already indicated.

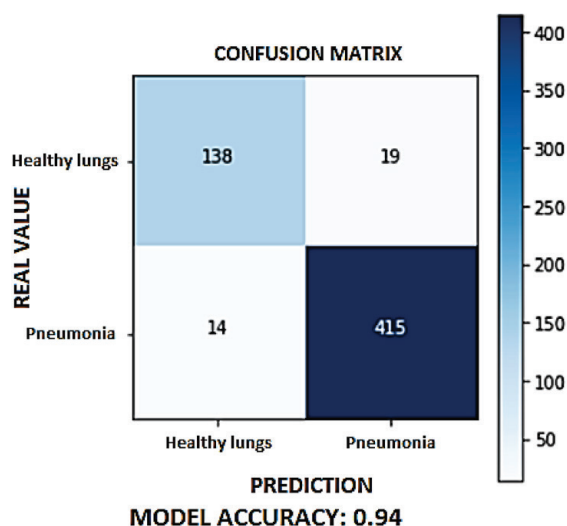


Figure 15 Model confusion matrix

4 CONCLUSION

The aim of this paper was to develop a model of an intelligent system that receives x-ray image of the lung as an input parameter and, based on the processed image, returns the possibility of pneumonia as an output. The implementation of the mentioned functionality was implemented using a transfer learning methodology based on already defined convolution neural network architectures.

In this example, the VGG16 architecture was used, consisting of a total of 16 layers, which greatly contributes to the accuracy of the model using previously learned principles for the classification of image data. After validation of the system, the model shows extremely good results on the validation dataset. For a more complete and qualitative prediction of pneumonia, more data should be available to make the model more representative for decision making and to assist radiologists in making diagnoses.

The model was later integrated with the technologies used to build web applications using the Flask framework. Creating a web application based on the implemented model contributes to the availability of the model, since web applications are available from anywhere at any time, with the requirement that they are connected to the internet. Intuitiveness and ease of use of the application are the main factors for widespread use, which would reduce the number of misdiagnosis of pneumonia.

Acknowledgements

This paper describes the results of research being carried out within the project "Centar održivog razvoja"/"Center of sustainable development", co-financed by the European

regional development fund and implemented within Operational Programme Competitiveness and Cohesion 2014 – 2020, based on the call "Investing in Organizational Reform and Infrastructure in the Research, Development and Innovation Sector".

5 REFERENCES

- [1] Image dataset, <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> (02.10.2019)
- [2] Hrnić, M. (2019). Design of detector model for RTG imaging, Polytechnic of Međimurje in Čakovec. <https://repozitorij.mev.hr/islandora/object/mev%3A1004> (18.09.2019)
- [3] Bishop, C. (2007). *Pattern Recognition and Machine Learning*. Springer, pp 738.
- [4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press, pp 787.
- [5] Nielsen, M. (2019). *Neural Networks and Deep Learning*. Online book.
- [6] Backpropagation algorithm, <https://towardsdatascience.com/understanding-backpropagation-algorithm-7bb3aa2f95fd> (23.09.2019)
- [7] Activation functions, <https://ieeexplore.ieee.org/document/8404724> (25.09.2019)
- [8] Williams, A. (2017). *Convolutional Neural Networks in Python*. CreateSpace Independent Publishing Platform, pp 140.
- [9] Transfer learning, <https://machinelearningmastery.com/transfer-learning-for-deep-learning/> (29.09.2019.)
- [10] Gulli, A. (2017). *Deep Learning with Keras*. Packt Publishing, pp 318.
- [11] Tensorboard, <https://www.tensorflow.org/tensorboard> (02.10.2019)
- [12] Confusion matrix, <https://ieeexplore.ieee.org/document/7326461> (02.10.2019)

Author's contacts:

Željko KNOK, mr. sc., senior lecturer
Polytechnic of Međimurje in Čakovec,
Bana Josipa Jelačića 22a, 40000 Čakovec, Croatia
zknok@mev.hr

Klaudio PAP, PhD, Prof.
University of Zagreb,
Faculty of Graphic Arts,
Getaldiceva 2, 10000 Zagreb, Croatia
kpap@grf.hr

Marko HRNČIĆ, student
Zagreb University of Applied Sciences
Mlinarska cesta 38, 10000 Zagreb, Croatia
mahmrcic@gmail.com

LOAD FREQUENCY CONTROL OF MULTI AREA INTERCONNECTED POWER SYSTEM USING DIFFERENTIAL EVOLUTION ALGORITHM

Alireza SINA, Damanjeet KAUR

Abstract: In this paper, Proportional Integral Derivative (PID) controller is designed using Differential Evolution (DE) algorithm to Load Frequency Control (LFC) in three areas of an interconnected power system. The proposed controller has appropriate dynamic response, so it increases damping in transient state in unhealthy conditions. Different generators have been used in three areas. Area 1 includes thermal non-reheat generator and two thermal reheat generators; area 2 includes hydro and thermal non-reheat generators, and area 3 includes hydro and thermal reheat generators. In order to evaluate the performance of the controller, Sim/Matlab software is used. Simulation results show that the controller designed using DE algorithm is not affected by load changes, disturbance, or system parameters changes. Comparing the results of proposed algorithm with other load frequency control algorithms, such as PSO and GA, it has been found that this method has a more appropriate response and satisfactory performance.

Keywords: ACE; DE; Frequency Deviation; LFC; Tie Line Power Deviation

1 INTRODUCTION

A modern power system has a number of generating units. These units are located in different areas. These areas are connected to each other with the help of a tie line to exchange power. When load change occurs in a modern power system, the balance between generated power and load is disturbed. This effect leads to a frequency deviation from the rated value, and the tie line power among different areas is changed from the scheduled value. If large deviations occur, power system may become unstable too.

For successful operation of power system, the balance between generated power and load demand should be maintained. In load frequency control (LFC) oscillations are damped and can maintain frequency and appropriate transient behaviour. Automatic generation control (AGC) includes load frequency control loop and an automatic voltage regulator (AVR) loop that regulates the frequency and power flow [1]. The main purpose of automatic generation control is to maintain frequency at rated value and tie line power at scheduled values when unusual conditions such as load changes, disturbance and system parameters change. For this purpose, intelligent control methods are developed to get the desired dynamic response.

In [2-7], intelligent techniques are used to optimize the PID / PI controller parameters. In [2], Laurent series is used to optimize the PID controller parameters. In [3], DE algorithm is used to design a 2-Dof PID (Parallel 2 degree freedom off) controller in for load frequency control in an interconnected power system. To eliminate frequency oscillations and tie line power, authors used SA (Simulated Annealing) algorithm for AVR [4]. In [5], direct synthesis method is used. In this method PID controller parameters are determined by linear algebraic equations. MUGA (Multi objective uniform diversity genetic algorithm) algorithm is implemented to optimize the PI/PID controller parameters [6]. In [8-14] PI/PID controller parameters are optimized using fuzzy logic in combination with intelligent techniques.

In [8] SAMA (Self-adaptive modified algorithm) algorithm along with fuzzy logic is used to obtain PI controller parameters. A type-2 fuzzy PID controller is designed via BBBC (Bing bang-big crunch) algorithm [9]. In this method, possible uncertainty in large-scale complex systems is considered and type-2 fuzzy sets are used. In order to optimize the PID controller parameters, GA, ANN and ANFIS algorithms are used [10]. Results show that the ANFIS algorithm has a more appropriate response. In [11], TBLO (Teaching learning based optimization) algorithm is used to obtain fuzzy PID controller parameters.

A hybrid combination of Neuro and Fuzzy is proposed as a controller to solve the Automatic Generation Control (AGC) problem in a restructured power system that operates under deregulation on the bilateral policy.

Recently evolutionary algorithms are used more often. Differential evolution (DE) is a method that repeats the attempt to find optimal points by differentiating a criterion. In classical optimization methods, such as gradient and Newton, a derivative operator is used, while in results of DE the quality criterion search method is used [15]. DE optimizes a problem by maintaining a population of candidate solutions and creating new candidate solutions by combining existing ones according to its simple formulae, and then keeping the candidate solution which has the best score or fitness on the optimization problem at hand. In this way, an optimization problem is treated as a black box that merely provides a measure of quality given by a solution, therefore the gradient is not needed.

In this paper, a PID controller is used for load frequency control in a three area interconnected power system. DE algorithm is used to optimize the controller parameters. By comparing the obtained results of simulations with other intelligent methods, including GA and PSO, the superiority of the proposed algorithm is specified. The impact of sudden load changes and system parameters changes is also considered in the proposed method.

2 THE SYSTEM MODEL

A three-area power system as shown in Fig. 1 is considered to evaluate the proposed method. Different generation sources are used in each area. Area 1 includes thermal non-reheat and reheat generators, while areas 2 and 3 include hydro and thermal reheat. System parameters are defined in Tab. 1. The connection between different areas is shown in Fig. 1. The dynamic models of generators used in the simulations are explained in [16-18].

As occurrence of any disturbance in each area causes frequency changes in all areas and tie line power deviation from scheduled values. A PID controller is used in each area in order to quickly damp oscillations and to reduce the overshoot and undershoot values. DE algorithm is used to optimize the parameters. Controller input is *ACE* signal. *ACE* is a linear combination of tie line power frequency that can be defined for each area *i*'th as Eq. (1).

$$ACE_i = \sum_j \Delta P_{tie,ij} + \beta_i \Delta f_i \quad (1)$$

In Eq. (1), $\Delta P_{tie,ij}$ is the power exchange between areas *i* and *j*, β_i is basic frequency coefficient of area *i* and Δf_i is frequency deviation of area *i*. In a multi-area power system, T_{ij} is considered as synchronizing coefficient between areas *i* and *j*, and Δp_i is also considered as load disturbances in area *i*.

3 DESIGN OF CONTROLLER

Differential Evolution (DE) technique is a population-based heuristic optimization algorithm [14-15]. Benefits of DE are: Simplesness, efficiency, and real coding. Generally, the base of DE is Genetic Algorithm (GA). However, unlike simple GA that uses binary coding for representing problem parameters, DE uses real coding of floating point numbers. The crucial idea behind DE is a scheme for generating trial parameter vectors. Basically, DE works with two populations; old generation and new generation of the same population. It works through a simple cycle of stages as Fig. 2.

Stage 1: Initial Population

A population of fixed size to generated where each variable has its lower and upper bounds.

Stage 2: Mutation

Selection of vectors to a base individual in order to explore the search space.

$$V_{iG+1} = X_{r1G} + F_1(X_{r2G} - X_{r3G}) \quad (2)$$

$$V_{iG+1} = X_{bestG} + F_1(X_{r2G} - X_{r3G}) \quad (3)$$

$$V_{iG+1} = X_{r1G} + F_1(X_{r2G} - X_{r3G}) + F_2(X_{bestG} - X_{r1G}) \quad (4)$$

Where: $i = 1, \dots, NP$; $r_1, r_2, r_3 \in \{1, \dots, NP\}$ are randomly selected and satisfy; $r_1 \neq r_2 \neq r_3 \neq i$; $F \in [0, 1]$, *F* is the control parameter/mutation factor proposed by Storn and Price [3].

Stage 3: Recombination or crossover

Mix successful solutions from the previous generation with current donors. The perturbed individual, $V_{iG+1} = (v_{1,iG+1}, \dots, v_{n,iG+1})$ and the current population member, $V_{iG} = (x_{1,iG}, \dots, x_{n,iG})$ are subject to the crossover operation, which finally generates the population of candidates, or "trial" vectors, $U_{iG+1} = (u_{1,iG+1}, \dots, u_{n,iG+1})$ as follows:

$$U_{i,jG+1} = \begin{cases} V_{i,jG+1} & \text{if } rand_{i,j} \leq CR \text{ or } j = I_{rand} \\ X_{i,jG} & \text{otherwise} \end{cases} \quad (5)$$

With $rand_{i,j} \sim U(0, 1)$, I_{rand} is a random integer from (1, 2, ..., *D*) where *D* is the solution's dimension, i.e. number of control variables and $CR \in [0, 1]$.

Stage 4: Selection

To keep the population size constant over subsequent generations, selection operation is performed. In this operation the target vector X_{iG} is compared with the trial vector V_{iG+1} and the one with the better fitness value is admitted to the next generation. The selection operation in DE can be represented by:

$$X_{iG+1} = \begin{cases} V_{iG+1} & \text{if } F(U_{iG+1}) < F(X_{iG}) \\ X_{iG} & \text{otherwise} \end{cases} \quad (6)$$

Proportional integral derivative controller (PID) is one of the most popular feedback controllers in the process industries because of its stability and fast response.

Design of PID controller requires determination of the three main parameters, Proportional gain K_p , Integral gain K_i and Derivative gain K_d . The integral square error (*ISE*) criterion is considered as the objective function for the present work which is described in Eq. (7).

$$ISE = \int_0^{t_{sim}} (\Delta f^2 + \Delta P_{tie}^2) dt \quad (7)$$

Where Δf is the system frequency deviation in area-*i* and area-*j*, respectively, ΔP_{tie}^2 is the incremental change in tie line power and t_{sim} is the time range of simulation. Fig. 3 shows flowchart of proposed DE and PID controller. The problem constraints are the controller parameter bounds. Therefore, the design problem can be formulated as the following optimization problem.

Minimum *ISE*. Subject to:

$K_p \min \leq K_p \leq K_p \max$, $K_i \min \leq K_i \leq K_i \max$ and $K_d \min \leq K_d \leq K_d \max$. K_{\min} , K_{\max} are the minimum and maximum value of the control parameters. The flowchart of the DE-PID algorithm is shown Fig. (3).

4 THE SIMULATION RESULTS

An interconnected power system includes three-area power system as shown in Fig. 1, which is simulated in Sim/Matlab software. System parameters have been defined according to the Tab. 1 and their values are shown in Tab. 2.

Power required in each area is considered to be $\Delta p_L = 0.1$ pu MW. PID controller is used in each area for automatic

generation control. The controller parameters are optimized using DE, PSO, and GA algorithms.

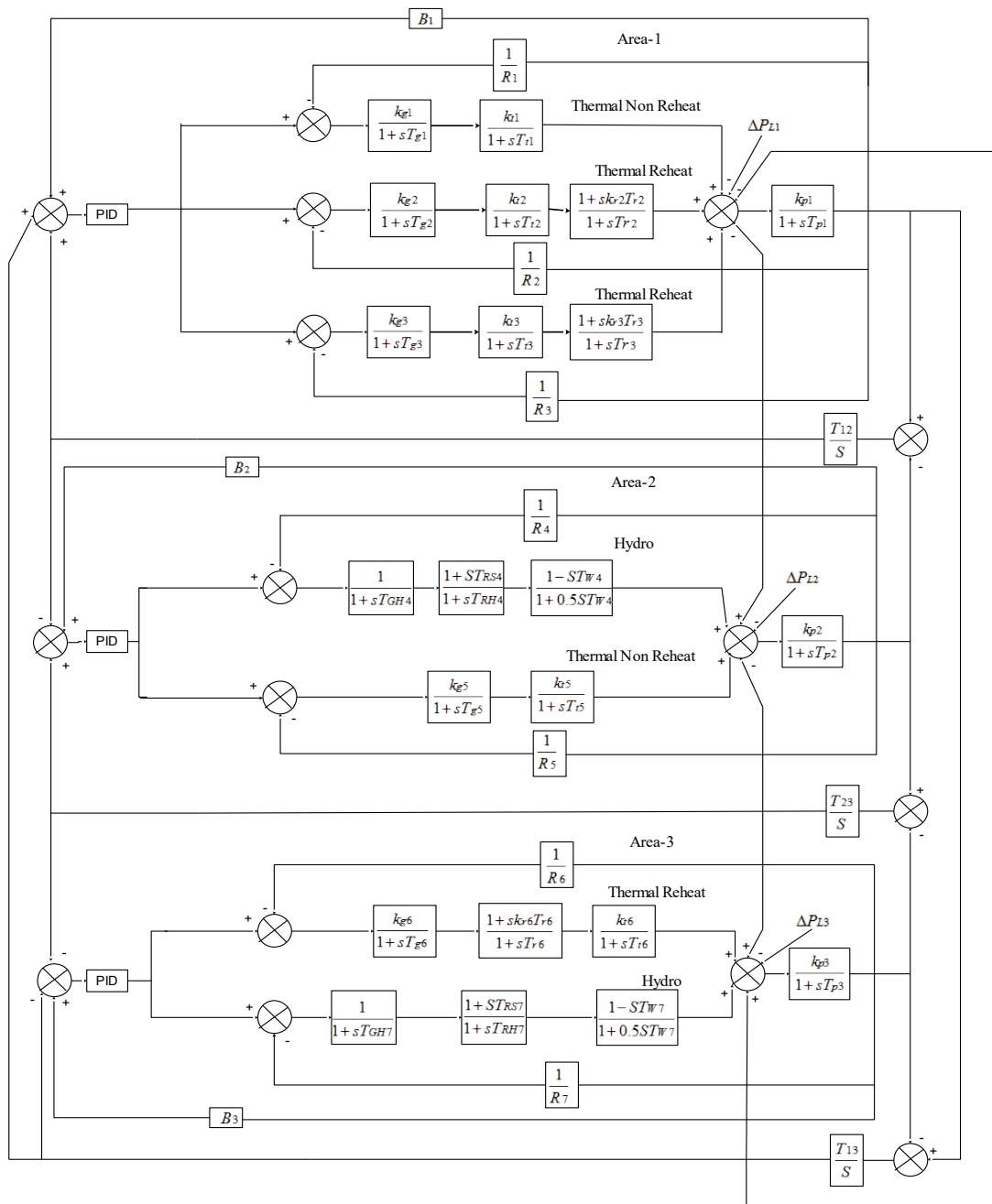


Figure 1 Three area multi source power system structure

Table 1 System parameters

Nomenclature			
ACE_i	Area control error of i^{th} area	R	Droop characteristic
F	Area frequency	ΔF_i	Incremental change in frequency of i^{th} control area
B_i	Frequency bias constant of i^{th} control area	TP	Subsystem equivalent time constant
T_g	Governer time constant	KP	Subsystem equivalent gain constant
T_t	Turbine time constant	T_{ij}	Synchronizing coefficient between area i and j
T_r	System turbine reheat time constant	$epfk$	Economic participation factor of k^{th} generation unit
TGH	Hydro turbine speed governer main servo time constant	$cpfk$	Constant participation factor between k^{th} GENCO and i^{th} DISCO
TRS	Hydro turbine speed governer reset time	$\Delta P_{tie,ij}$	Scheduled power tie line power flow between area i and j
TRH	Hydro turbine speed governer transient droop time constant	ΔPD_i	Total power control and uncontrolled for DISCO area i^{th}
TW	Nominal starting time of water in penstock	ΔP_{gk}	Power output of k^{th} generating unit

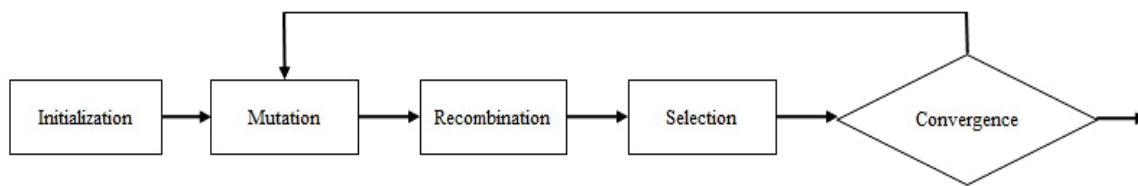


Figure 2 Simple cycle of DE

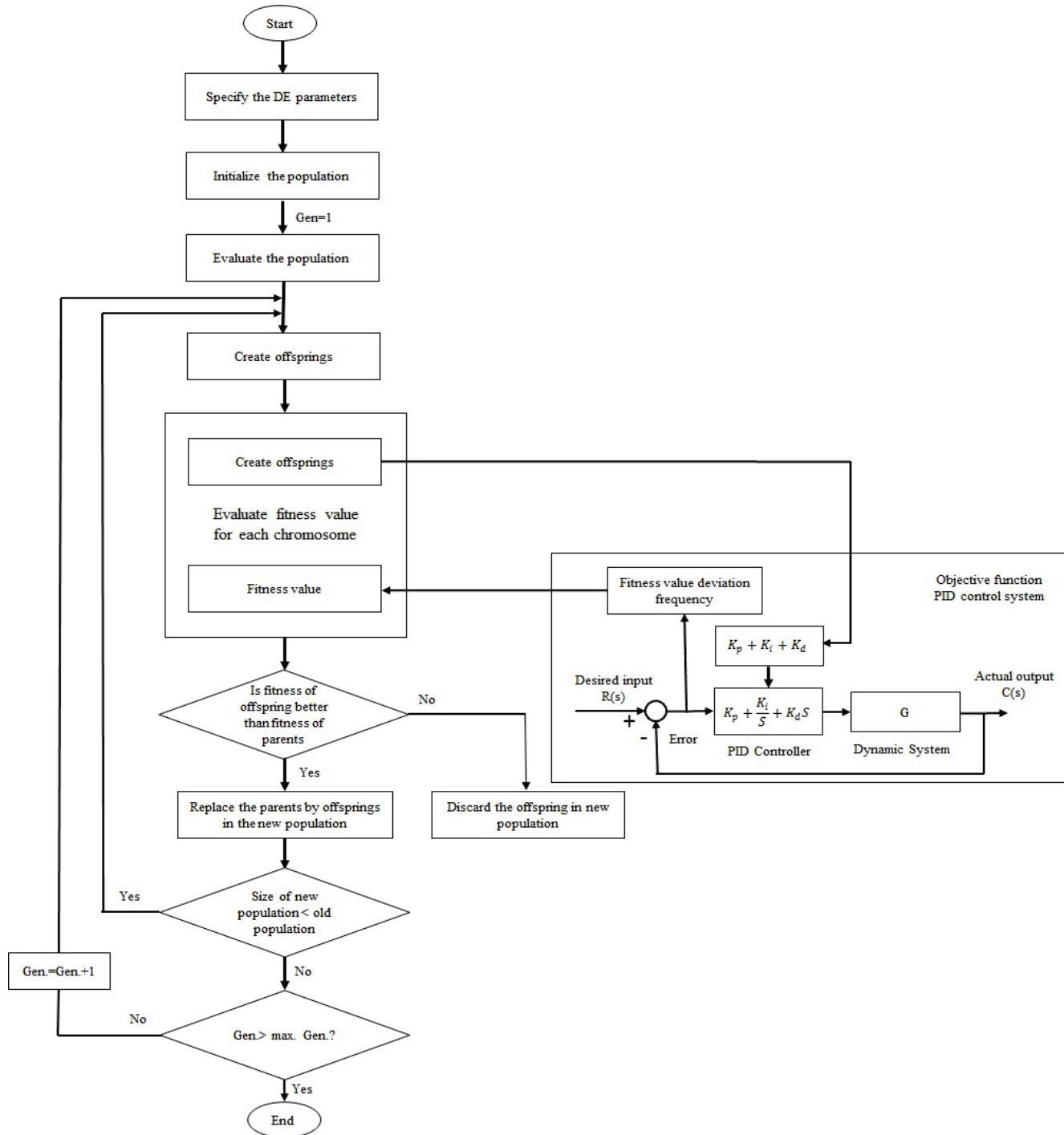


Figure 3 The flowchart of the DE-PID algorithm

Figs. 4 and 5 show the dynamic response of the system frequency deviation and tie-line power exchange between different areas respectively, in present of DE, PSO and GA

algorithms. From Figs. 4 and 5 it is clear that overshoot and undershoot is minimal.

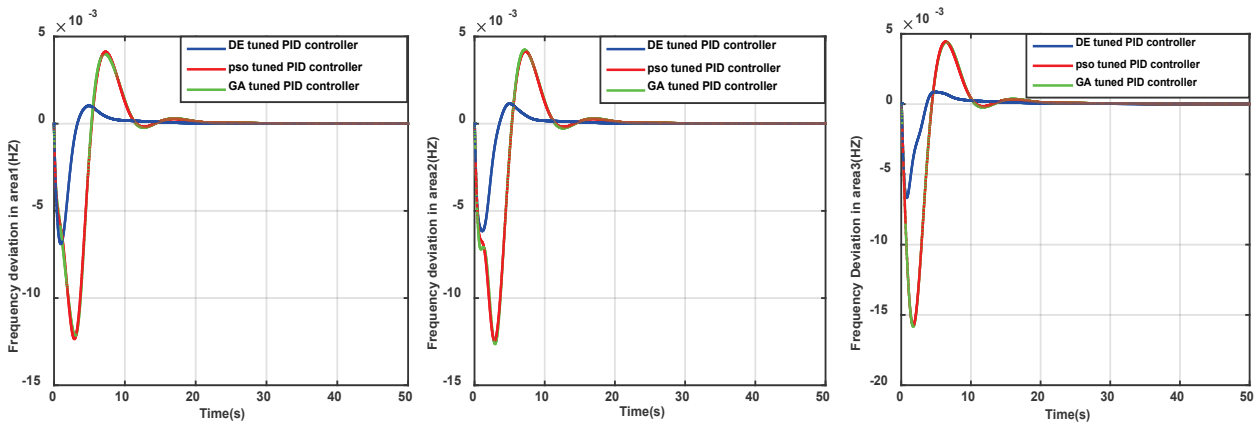


Figure 4 Dynamic response of the system frequency deviation

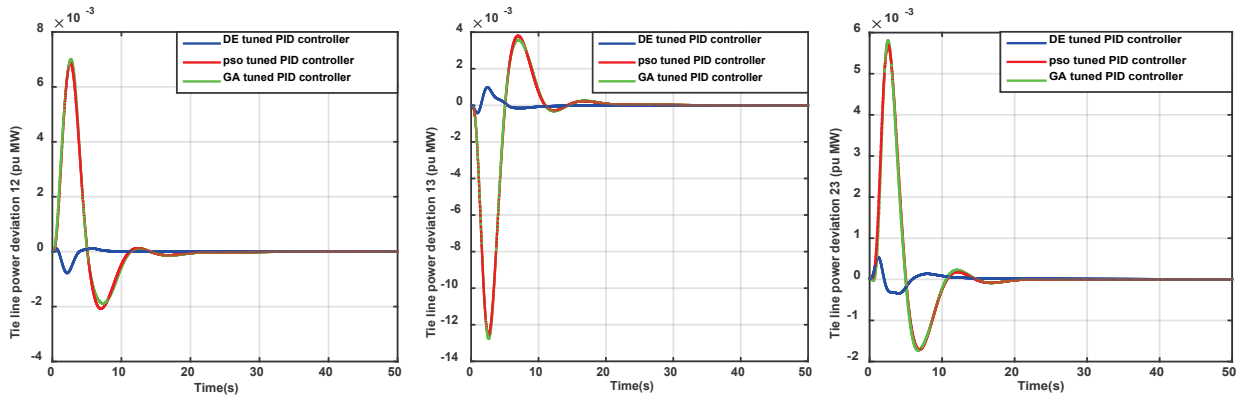


Figure 5 Dynamic response of tie line power deviation

Table 2 Power system parameter values

Area 1:	Area 2:	Area 3:
$T_{g1} = 0.08$ s	$T_{GH4} = 0.1$ s	$T_{g6} = 0.07$ s
$T_{t1} = 0.35$ s	$T_{RH4} = 10$ s	$T_{t6} = 0.36$ s
$R_1 = 0.3333 \frac{\text{Hz}}{\text{pu MW}}$	$T_{RS4} = 0.513$ s	$K_{r6} = 0.33$
$k_{g1} = 1$	$T_{W4} = 1$ s	$T_{r6} = 10$ s
$k_{t1} = 1$	$R_4 = 0.32 \frac{\text{Hz}}{\text{pu MW}}$	$R_6 = 0.289 \frac{\text{Hz}}{\text{pu MW}}$
$T_{g2} = 0.875$ s	$k_{g4} = k_{t4} = 1$	$k_{g6} = k_{r6} = 1$
$T_{t2} = 0.375$ s	$T_{g5} = 0.075$ s	$T_{GH7} = 0.015$ s
$k_{r2} = 0.3113$	$T_{t5} = 0.38$ s	$T_{RH7} = 8.75$ s
$T_{r2} = 10.6$ s	$R_5 = 0.2963 \frac{\text{Hz}}{\text{pu MW}}$	$T_{RS7} = 0.1513$ s
$R_2 = 0.32 \frac{\text{Hz}}{\text{pu MW}}$	$k_{g5} = k_{t5} = 1$	$T_{W7} = 1.5$ s
$k_{g2} = k_{t2} = 1$	$k_{p2} = 20 \frac{\text{Hz}}{\text{pu MW}}$	$R_7 = 0.3077 \frac{\text{Hz}}{\text{pu MW}}$
$T_{g3} = 0.08$ s	$T_{p2} = 120$ s	$k_{g7} = k_{t7} = 1$
$T_{t3} = 0.3$ s	$B_1 = 0.495 \frac{\text{pu}}{\text{MW Hz}}$	$k_{p3} = 20 \frac{\text{Hz}}{\text{pu MW}}$
$k_{r3} = 0.5$	$f = 60$ Hz	$T_{p3} = 120$ s
$T_{r3} = 10$ s		$B_3 = 0.542 \frac{\text{pu}}{\text{MW Hz}}$
$R_3 = 0.33 \frac{\text{Hz}}{\text{pu MW}}$		$f = 60$ Hz
$k_{g3} = k_{t3} = 1$		$T_{t2} = T_{t3} = T_{t13} = 0.543$ pu/H
$k_{p1} = 20 \frac{\text{Hz}}{\text{pu MW}}$		
$T_{p1} = 120$ s		
$B_1 = 0.532 \frac{\text{pu}}{\text{MW Hz}}$		
$f = 60$ Hz		

Tab. 3 shows the amount of overshoot, undershoot and settling time in exchange of power system all three algorithms. The results show that the dynamic performance of DE algorithm is suitable.

4.1 Controller Performance against Load Changes

In order to evaluate the performance of the controller against possible load variation, the load of all areas changes from -50% to $+50\%$. Figs. 6 and 7 show the frequency deviation and tie-line power exchange between different areas respectively. The amount of overshoot, undershoot and settling time in various states of load changes for three algorithms is shown in Tabs. 4 and 5.

Results in Tabs. 4 and 5 clearly show that increment in rate of load, increases the measured values and the DE algorithm has more satisfactory performance in the case of load changes. The effect of reducing the amount of load up to -50% , settling time for the tie-line power exchange between different areas will be zero.

This shows that the controller proposed is well managed from the beginning and control deviations without any delay within the desired precision and show satisfactory performance. In the case of load reduction up to -25% , the settling time for Δp_{tie23} also has a similar situation.

Table 3 Overshoot, undershoot and settling time for base case with PID controller using DE, PSO and GA algorithms

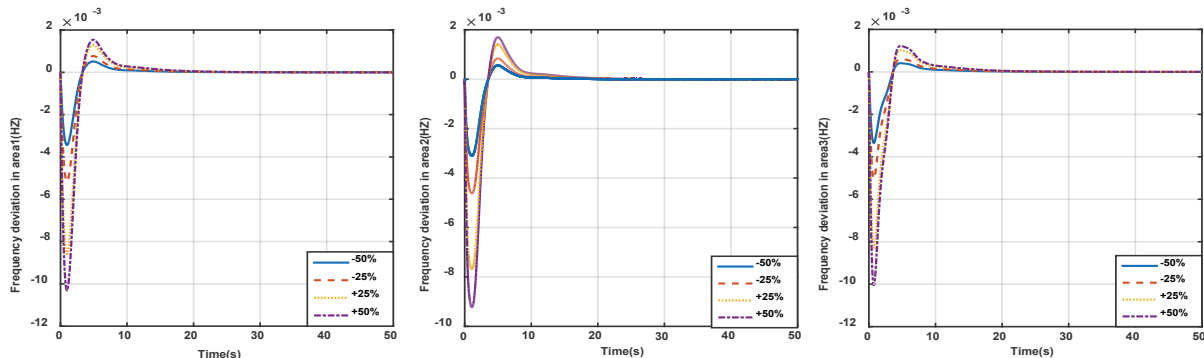
Parameters	DE			PSO			GA		
	OS	US	ST	OS	US	ST	OS	US	ST
Δf_1	0.00102	-0.0068	6.96	0.0041	-0.0122	10.79	0.004	-0.0121	10.82
Δf_2	0.00115	-0.00614	7.209	0.00415	-0.0124	10.75	0.0042	-0.0126	10.75
Δf_3	0.000818	-0.0067	7.09	0.0043	-0.0157	9.84	0.0043	-0.0158	9.8
ΔP_{tie12}	0.000113	-0.00077	2.83	0.0068	-0.00207	9.95	0.007	-0.00188	10.16
ΔP_{tie13}	0.00097	-0.00041	3.44	0.0037	-0.0125	10.28	0.00356	-0.0127	10.35
ΔP_{tie23}	0.00051	-0.00036	1.35	0.00567	-0.00172	9.523	0.00578	-0.00174	9.51

Table 4 Overshoot, undershoot and settling time of Δf_1 , Δf_2 , Δf_3 for different loading conditions

Controller	Changes in load (%)	Δf_1			Δf_2			Δf_3		
		OS	US	ST	OS	US	ST	OS	US	ST
DE	-50%	0.00051	-0.0034	5.23	0.000574	-0.00307	5.67	0.000409	-0.0033	3.17
	-25%	0.00076	-0.00515	6.37	0.00086	-0.0046	6.63	0.00061	-0.00502	6.37
	+25%	0.00128	-0.0085	7.43	0.00143	-0.0076	7.61	0.00102	-0.0083	7.54
	+50%	0.00154	-0.0103	7.83	0.00172	-0.00921	7.92	0.00122	-0.01005	8.01
PSO	-50%	0.00206	-0.0061	10.28	0.00207	-0.00621	10.2	0.00219	-0.00789	9.33
	-25%	0.00309	-0.00924	10.6	0.00311	-0.00932	10.53	0.00328	-0.01184	9.62
	+25%	0.0051	-0.01541	10.93	0.005196	-0.0155	10.87	0.005475	-0.01973	9.95
	+50%	0.00617	-0.01849	11.03	0.006236	-0.01865	10.96	0.00657	-0.02368	10.05
GA	-50%	0.002	-0.00608	10.33	0.00211	-0.0063	10.24	0.002175	-0.00793	9.36
	-25%	0.003	-0.0091	10.65	0.00317	-0.0094	10.57	0.003263	-0.0119	9.65
	+25%	0.005	-0.01518	10.96	0.005285	-0.01577	10.87	0.005438	-0.01938	9.92
	+50%	0.006	-0.01825	11.03	0.006342	-0.01892	10.96	0.006526	-0.02379	10.02

Table 5 Overshoot, undershoot and settling time of ΔP_{tie12} , ΔP_{tie13} and ΔP_{tie23} for different loading conditions

Controller	Changes in load (%)	ΔP_{tie12}			ΔP_{tie13}			ΔP_{tie23}		
		OS	US	ST	OS	US	ST	OS	US	ST
DE	-50%	0.000056	-0.00038	0	0.00048	-0.00020	0	0.000259	-0.00018	0
	-25%	0.000085	-0.00057	2.58	0.00073	-0.00031	3.14	0.00038	-0.00027	0
	+25%	0.00014	-0.00096	3.03	0.001222	-0.00052	3.76	0.00064	-0.00045	1.52
	+50%	0.00017	-0.00116	3.14	0.00146	-0.00062	4.08	0.00077	-0.00054	4.4
PSO	-50%	0.00343	-0.00103	9.17	0.001898	-0.00626	9.76	0.002836	-0.00086	8.66
	-25%	0.00515	-0.00155	9.67	0.002847	-0.0094	10.1	0.004254	-0.00129	9.23
	+25%	0.00858	-0.00259	10.14	0.004745	-0.01567	10.4	0.00709	-0.00215	9.73
	+50%	0.0103	-0.0031	10.26	0.00569	-0.0188	10.49	0.008507	-0.00258	9.85
GA	-50%	0.0035	-0.00094	9.35	0.001784	-0.00637	9.83	0.002894	-0.00087	8.66
	-25%	0.0052	-0.0014	9.86	0.002676	-0.00957	10.15	0.004341	-0.0013	9.23
	+25%	0.00875	-0.00235	10.33	0.004459	-0.01595	10.45	0.007235	-0.00217	9.69
	+50%	0.0105	-0.00282	10.45	0.005351	-0.01914	10.52	0.008682	-0.00261	9.81

**Figure 6** System response of frequency deviation for different loading changes

4.2 Evaluation of Controller against Changes in System Parameters

In order to further evaluate the performance of the proposed controller under various conditions, several system parameters such as T_t (turbine time constant), T_r (reheat time constant), T_{GH} (hydro turbine speed governor main servo time constant), T_{RS} (hydro turbine speed governor reset time), T_{RH} (hydro turbine speed governor transient droop

time constant) have been changed in the range of -25% to $+25\%$ compared to their nominal values by considering the coefficients of an optimal controller.

Important dynamic indices are overshoot, undershoot and settling time which are measured and tabulated in Tabs. 6 and 7. The results show that the values of the indices measured are in an acceptable range and close to the values of the parameters of the system.

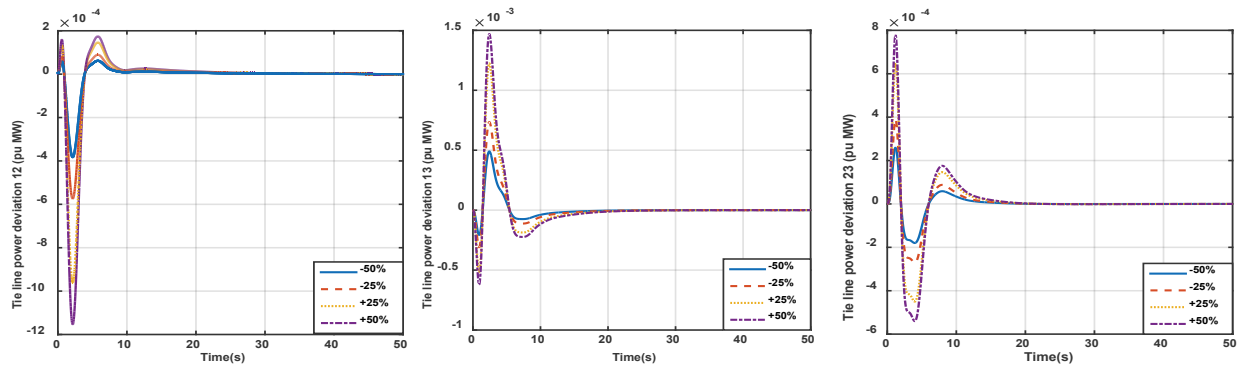


Figure 7 System response of tie-line power deviation for different loading changes

Table 6 Overshoot, undershoot and setting time of Δf_1 , Δf_2 , Δf_3 for different values of system parameters

System parameters	Changes in load (%)	Δf_1			Δf_2			Δf_3		
		OS	US	ST	OS	US	ST	OS	US	ST
T_i	-25%	0.00099	-0.00629	7.3	0.00108	-0.0057	7.46	0.00081	-0.0059	7.3
	+25%	0.0011	-0.0074	6.3	0.00147	-0.00659	7.61	0.00156	-0.0075	7.21
T_r	-25%	0.001	-0.0068	6.65	0.00117	-0.00611	7.03	0.0008	-0.00659	6.94
	+25%	0.00102	-0.0069	7.14	0.00113	-0.00616	7.31	-0.0082	-0.0067	7.17
T_{GH}	-25%	0.00102	-0.00688	6.93	0.00114	-0.00612	7.19	0.00081	-0.0067	7.07
	+25%	0.00103	-0.0068	6.99	0.00115	-0.0061	7.22	0.00082	-0.0066	7.11
T_{RS}	-25%	0.00104	-0.00682	7.03	0.00116	-0.00617	7.23	0.00085	-0.0066	7.15
	+25%	0.0010	-0.0069	6.89	0.00113	-0.00611	7.18	0.00077	-0.0067	7.01
T_{RH}	-25%	0.00106	-0.00707	6.38	0.00126	-0.0063	6.43	0.00113	-0.0073	7.20
	+25%	0.00090	-0.00677	7.31	0.00105	-0.00606	7.51	0.00082	-0.0063	7.28

Table 7 Overshoot, undershoot and setting time of $\Delta P_{tie 12}$, $\Delta P_{tie 13}$ and $\Delta P_{tie 23}$ for different values of system parameters

System parameters	Changes in load (%)	$\Delta P_{tie 12}$			$\Delta P_{tie 13}$			$\Delta P_{tie 23}$		
		OS	US	ST	OS	US	ST	OS	US	ST
T_i	-25%	8.01×10^{-5}	-0.00055	2.57	0.00070	-0.00027	3.93	0.00043	-0.00039	0
	+25%	0.00028	-0.00109	2.89	0.00154	-0.00058	3.1	0.00062	-0.00051	2.66
T_r	-25%	0.000175	-0.00083	2.80	0.00112	-0.00037	3.30	0.00046	-0.00039	0
	+25%	0.000105	-0.000726	2.89	0.00088	-0.00044	3.53	0.00055	-0.00033	1.46
T_{GH}	-25%	0.000116	-0.00078	2.85	0.00098	-0.00042	3.44	0.00052	-0.00036	1.36
	+25%	0.0001107	-0.000763	2.86	0.00097	-0.00041	3.47	0.00051	-0.00035	1.31
T_{RS}	-25%	9.47×10^{-5}	-0.000689	2.82	0.00092	-0.00037	3.59	0.00051	-0.00038	1.31
	+25%	0.000144	-0.000859	2.89	0.00104	-0.00046	3.32	0.00053	-0.00036	1.35
T_{RH}	-25%	0.0002521	-0.00126	3.27	0.00180	-0.00077	3.44	0.00066	-0.00055	2.75
	+25%	0.000113	-0.00064	2.41	0.00066	-0.00025	2.94	0.00045	-0.00038	0

5 CONCLUSION

In this paper, the design of an intelligent controller in interconnected large power system has been presented. An extensive analysis of proposed LFC system controller in the interconnected power system has been done when execution of the changes load demand and changes in system parameters are taken into account. Results clearly show that PID controller reaches to zero deviations in steady state in frequency and tie line power quickly.

In order to show superiority of PID designed controller using DE algorithm, the results of the Differential Evolution technique of parameters such as frequency and power tie line changes compared with the results of the PSO and GA algorithms. Results prove that DE algorithm has better overshoot, undershoot and settling time in all possible as compared to PSO and GA. Furthermore, the results of simulations show that the proposed algorithm is not affected by changes in load and uncertainty in the system parameters and the controller has satisfactory dynamic operation.

6 REFERENCES

- [1] Prakash, S. & Sinha, S. (2014). Simulation based neuro-fuzzy hybrid intelligent PI control approach in four-area load frequency control of interconnected power system. *Applied Soft Computing*, 23, 152-164. <https://doi.org/10.1016/j.asoc.2014.05.020>
- [2] Padhan, D. G. & Majhi, S. (2013). A new control scheme for PID load frequency controller of single-area and multi-area power systems. *ISA transactions*, 52(2), 242-251. <https://doi.org/10.1016/j.isatra.2012.10.003>
- [3] Sahu, R. K., Panda, S., & Rout, U. K. (2013). DE optimized parallel 2-DOF PID controller for load frequency control of power system with governor dead-band nonlinearity. *International Journal of Electrical Power & Energy Systems*, 49, 19-33. <https://doi.org/10.1016/j.ijepes.2012.12.009>
- [4] Chandrakala, K. V. & Balamurugan, S. (2016). Simulated annealing based optimal frequency and terminal voltage control of multi source multi area system. *International Journal of Electrical Power & Energy Systems*, 78, 823-829. <https://doi.org/10.1016/j.ijepes.2015.12.026>
- [5] Anwar, M. N. & Pan, S. (2015). A new PID load frequency controller design method in frequency domain through direct

- synthesis approach. *International Journal of Electrical Power & Energy Systems*, 67, 560-569.
<https://doi.org/10.1016/j.ijepes.2014.12.024>
- [6] Nikmanesh, E., Hariri, O., Shams, H., & Fasihozaman, M. (2016). Pareto design of load frequency control for interconnected power systems based on multi-objective uniform diversity genetic algorithm (muga). *International Journal of Electrical Power & Energy Systems*, 80, 333-346.
<https://doi.org/10.1016/j.ijepes.2016.01.042>
- [7] Hota, P. & Mohanty, B. (2016). Automatic generation control of multi source power generation under deregulated environment. *International Journal of Electrical Power & Energy Systems*, 75, 205-214.
<https://doi.org/10.1016/j.ijepes.2015.09.003>
- [8] Khooban, M. H. & Niknam, T. (2015). A new intelligent online fuzzy tuning approach for multi-area load frequency control: Self Adaptive Modified Bat Algorithm. *International Journal of Electrical Power & Energy Systems*, 71, 254-261.
<https://doi.org/10.1016/j.ijepes.2015.03.017>
- [9] Yesil, E. (2014). Interval type-2 fuzzy PID load frequency controller using Big Bang–Big Crunch optimization. *Applied Soft Computing*, 15, 100-112.
<https://doi.org/10.1016/j.asoc.2013.10.031>
- [10] Mosaad, M. I. & Salem, F. (2014). LFC based adaptive PID controller using ANN and ANFIS techniques. *Journal of Electrical Systems and Information Technology*, 1(3), 212-222.
<https://doi.org/10.1016/j.jesit.2014.12.004>
- [11] Nayak, J. R., Pati, T. K., Sahu, B. K., & Kar, S. K. (2015). Fuzzy-PID controller optimized TLBO algorithm on automatic generation control of a two-area interconnected power system. Paper presented at the *Circuit, Power and Computing Technologies (ICCPCT 2015)*, *International Conference on*.
<https://doi.org/10.1109/ICCPCT.2015.7159427>
- [12] Sekhar, G. C., Sahu, R. K., Baliarsingh, A., & Panda, S. (2016). Load frequency control of power system under deregulated environment using optimal firefly algorithm. *International Journal of Electrical Power & Energy Systems*, 74, 195-211.
<https://doi.org/10.1016/j.ijepes.2015.07.025>
- [13] Dhillon, S. S., Lather, J. S., & Marwaha, S. (2015). Multi area load frequency control using particle swarm optimization and fuzzy rules. *Procedia Computer Science*, 57, 460-472.
<https://doi.org/10.1016/j.procs.2015.07.363>
- [14] Sahu, R. K., Sekhar, G. C., & Panda, S. (2015). DE optimized fuzzy PID controller with derivative filter for LFC of multi source power system in deregulated environment. *Ain Shams Engineering Journal*, 6(2), 511-530.
<https://doi.org/10.1016/j.asej.2014.12.009>
- [15] Rocca, P., Oliveri, G., & Massa, A. (2011). Differential evolution as applied to electromagnetics. *IEEE Antennas and Propagation Magazine*, 53(1), 38-49.
<https://doi.org/10.1109/MAP.2011.5773566>
- [16] Parmar, K. S., Majhi, S., & Kothari, D. (2012). Load frequency control of a realistic power system with multi-source power generation. *International Journal of Electrical Power & Energy Systems*, 42(1), 426-433.
<https://doi.org/10.1016/j.ijepes.2012.04.040>
- [17] Report, I. (1973). Dynamic models for steam and hydro turbines in power system studies. *IEEE Transactions on Power Apparatus and Systems* (6), 1904-1915.
<https://doi.org/10.1109/TPAS.1973.293570>
- [18] De Mello, F. & Ahner, D. (1994). Dynamic models for combined cycle plants in power system studies. *IEEE Transactions on Power Systems*, 9(3).
<https://doi.org/10.1109/59.336085>

Authors' contacts:

Alireza SINA, Faculty member of ACECR, Research scholar in UIET
 Panjab University,
 Sector 14, Chandigarh, Punjab 160014, India
sina@acecr.ac.ir

Damanjeet KAUR, Assistant Professor of Electrical & Electronics Engineering, UIET
 Panjab University,
 Sector 14, Chandigarh, Punjab 160014, India
djkb14@rediffmail.com

DESIGNING PLASTER MORTAR FOR A HIGH-RISE BUILDING

Zeljko KOS, Valentyn PARUTA, Olga GNYP, Leonid LAVRENYUK, Iryna GRYNKOVA

Abstract: The article rethinks the approach to the development of compositions of plaster mortar as the load on the plaster coating located on the 1st or 24th floor of the building differs significantly. To solve the problem within the framework of existing concepts, principles and regulatory requirements are not possible, as they are applicable to the solutions used in low-rise building. To achieve this goal, it is necessary to develop the scientific basis for the design of plaster mortar. In order to do that, it is necessary to analyze the processes occurring in the plaster coating when it is applied and hardened, the knowledge of the mechanism of the destruction of the "masonry - plaster coating" system, and the calculation and evaluation of stresses. The designated physical and mechanical parameters and the selected plaster mortar compositions should ensure the "work" of the system at the maximum level of such stresses.

Keywords: calculation of stresses; high-rise building; mechanical parameters; physical parameters; plaster mortar

1 INTRODUCTION

An increase in the share of high-rise buildings, together with the widespread use of new wall materials, requires a rethinking of the approach to the design of plaster mortar compositions. This is due to the fact that the impacts and loads on the plaster coating located on the 24th floor of a building are significantly different from those on the 1st, 2nd and 3rd floor (Fig. 1).

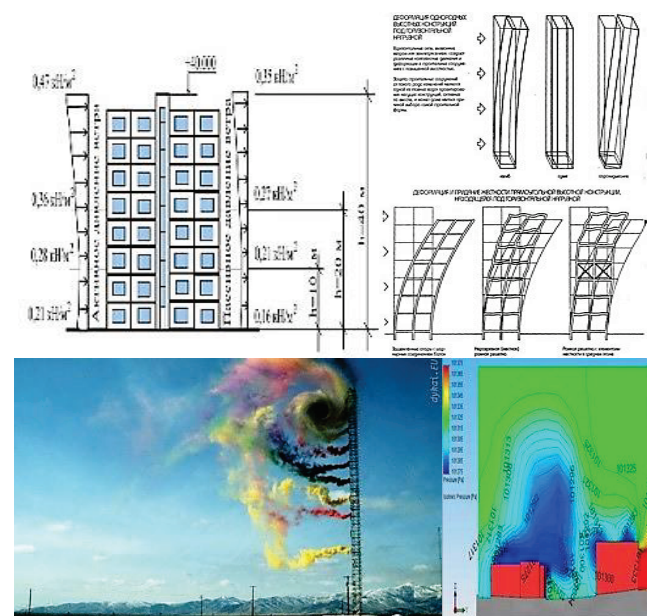


Figure 1 Variation of impacts and loads on the plaster coating with an increasing building height

It is also necessary to take into account that when building high-rise buildings on the territory of the Ukraine, the average density of autoclaved aerated concrete is mainly 150-600 kg/m³. Its properties (compressive strength, modulus of elasticity, temperature deformation) depend on the average density and significantly differ from the properties of traditional materials from which the walls are

built (ceramic and silicate bricks, blocks of lightweight concrete and rocks).

In order to prevent the destruction of plaster coating, it is necessary to calculate the values of its physical-mechanical characteristics (compressive and bending strength, elastic modulus, etc.), taking into account the stresses in it and the contact zone with the cladding that arises when the solution is hardened and the wall structure is deformed. In this case, the maximum permissible strain should be less than destructive. Selecting the composition of the solution needs to be carried out by taking into account the processes occurring during the application and hardening of the mortar to the masonry. It is necessary to increase the water-retaining capacity of the mixture, to reduce the shrinkage of plaster coating during hardening, to reduce the number of cracks that occur during hardening and to prevent or slow down their development.

Normative requirements for plaster mortars for autoclaved aerated concrete walls are also not taken into account as they are contradictory and, in our opinion, not substantiated. For example, compressive strength should be 1,5-7,5 MPa (Russia), 2,5 MPa (Ukraine) and 10 MPa (Germany). The flexural strength should be 1-1,25 MPa (Ukraine) and 2 MPa (Russia). The value of adhesion to aerated concrete laying should be 0,15-3 MPa (Russia) and 0,5 MPa (Ukraine). The design principles of plaster mortar compositions proposed by the authors are new directions in the building materials' science [1].

The aim of the research was to obtain plaster mortar with high crack resistance, while ensuring the requirements of normative documents for medium density, compressive strength and bending, and other parameters.

To achieve the required goals, the development of new scientific bases for the design of plaster mortar compositions is required. It is necessary to analyze the processes occurring in the plaster coating when it is applied and hardened, to know the destruction mechanism of the "masonry - plaster coating" system calculation and calculation of stresses.

Recent works dealing with the subject matter indicate that the durability of the wall structure depends to a large

extent on the number of defects in the plaster coating and the contact area between it and the masonry [2, 8, 9]. The destruction of the "masonry - plaster coating" system is due to the accumulation and development of micro- and macro-cracks in its structure. To assess the resource of such a system, it is necessary to determine the internal and external factors, the degree of their impact, the calculation of the stress state, knowledge of the processes of nucleation, accumulation of damages and growth of macrocracks. The assigned physical and mechanical parameters and compositions of plaster mortar should ensure the "work" of the system at the maximum level of such stresses.

2 ANALYSIS OF THE PROCESSES OCCURRING IN PLASTER MORTAR

Masonry made of autoclaved aerated concrete has a high capillary potential due to considerable pore volume (520 mm³/g) and their high specific surface area (22-34 m²/g) [2]. After the mortar is applied to masonry, because of its low water-retaining capacity, liquid is sucked out from the lyophilic pores and capillaries of the masonry material (Fig. 2a). The pore filling rate (v) is determined by the Poiseuille equation:

$$v = -\frac{r^2}{8\eta} \frac{\Delta p}{l}, \quad (1)$$

where: l - is the length of the area of the absorbed liquid, η - is its viscosity, Δp - the pressure drop across section l is equal to the capillary pressure of the meniscus:

$$\Delta p = -2\sigma_{12} \cos \frac{q}{r}, \quad (2)$$

Therefore, the dissolution of cement and the formation of a supersaturated solution occur with reduced water content. Because of this, the incomplete hydration of cement occurs and non-equilibrium, metastable neoplasms of the following type are formed:

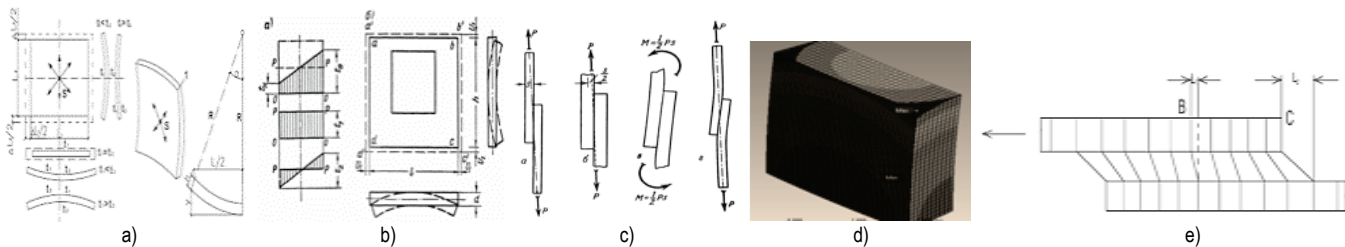
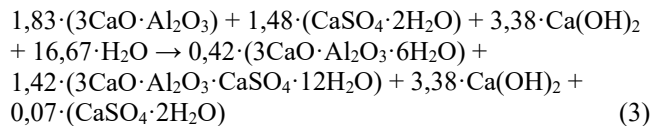


Figure 3 Temperature and humidity deformation of the masonry (a, b) and the stresses in it (d)

$$\Delta L_p = \alpha \cdot \Delta T_p \cdot L, \quad (5)$$

The loss of water leads to the shrinkage of the plaster mortar – 2,5 - 5,8 mm/m [3]. Since the aerated concrete masonry "holds back" these deformations, this leads to stresses in it (δ), which are seven times higher than its tensile strength:

$$\delta = \frac{\Delta \varepsilon^* \cdot E}{1 - \mu}, \quad (4)$$

where: E and μ are the modulus of elasticity and the Poisson's ratio of the plaster coating; $\Delta \varepsilon^*$ - the difference of deformations of plaster and aerated concrete base [6].

Because of these stresses and the fact that a decrease in the degree of hydration of the cement has led to a decrease in the ultimate extensibility of the material (by 20 - 50%) [7, 9, 10], cracks develop in the plaster coating (on the surface and in the volume of the material), as well as in the contact area between the ceiling and masonry (Fig. 2c, 2d, 2e).

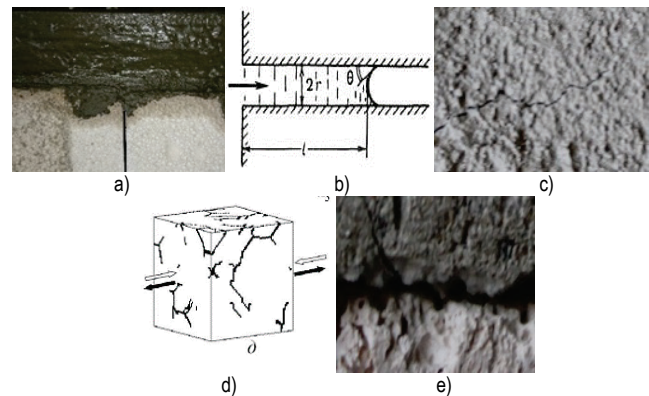


Figure 2 Low water retention solution on the masonry material (a), cracking in the plaster coating and the contact area between the ceiling and masonry (c, d, e)

During operation, they "develop" and are combined into trunk lines. The causes of cracks are temperature and humidity, deformation of plaster and masonry [4, 5] and the difference between them (Fig. 3c, 3d), the strain caused by them (Fig. 3a), moisture, ice, and corrosive materials.

Cyclic temperature-humidity effects lead to deformation (ΔL_p , ΔL_c , ΔL_{TB}) and additional stresses in the plaster coating (δ), which are the cause of the development of a whole family of main cracks.

$$\Delta L_c = \alpha \cdot \Delta T_c \cdot L, \quad (6)$$

$$\Delta L_{TB} = (\alpha \cdot \Delta L \cdot T \cdot F_m) + (\Delta L_q \cdot L), \quad (7)$$

where: $\Delta L_p, \Delta L_c$ - deformation of tension and compression; α - coefficient of temperature elongation of the material; $\Delta T_p, \Delta T_c$ - the temperature difference; L - the length of the wall element; ΔL_{TB} - heat and humidity deformation; F_m - an indicator of material defects; ΔL_q - humidity deformations [4].

$$\delta = \kappa \cdot \sqrt{[\alpha_t + b(L - c)] \delta^*}, \quad (8)$$

$$\delta^* = E \cdot \alpha_t \cdot \Delta t, \quad (9)$$

where: E - modulus of elasticity; α_t - coefficient of linear expansion of plaster coating; Δt - calculated temperature drop.

Masonry and plaster coating have different magnitudes of thermal deformations. The magnitude of these deformations is determined from the expression [6]:

$$\Delta L = L_0 \cdot \alpha_t \cdot \Delta t, \quad (10)$$

$$\Delta t = t_2 - t_1, \quad (11)$$

where: ΔL - the elongation or contraction of the wall structure is relative; L_0 the length of the wall structure at the time of construction; α_t - coefficient of thermal expansion and autoclaved aerated plaster coating [2, 6], Δt - changing the temperature of the wall structure; t_1 - environmental temperature at the time of construction of gas concrete masonry and applying stucco coatings; t_2 - the maximum and minimum temperature, which affect the wall structure in the summer and winter periods;

With temperatures as low as -20°C , the total compression deformation of an 8 m long wall, produced in summer at a temperature of $+30^\circ\text{C}$, with a coefficient of temperature expansion of aerated concrete laying of 8×10^{-6} grad-1 and a temperature change from $+30$ to -20°C , $e = 50^\circ\text{C}$ is 3.2 mm. In the summer, when heated to $+80^\circ\text{C}$ [2], the expansion deformation is 3,2 mm.

In the winter, the compression deformation of the plaster mortar (1: 4) will be 0,55 mm/m, and the total compression deformation of the plaster coating for the 8 m long wall will be 4,4 mm. In the summer, the total deformation expansion of the plaster coating will be 4,4 mm. Deformations of expansion and contraction cause tension (σ) in the masonry and the plaster covering which can be determined by converting the equation:

$$\Delta L : L_0 = \sigma : E, \quad (12)$$

where: ΔL - the elongation or contraction of the wall structure; L_0 - length of the wall structure at the time of erection; σ - stresses in N/mm^2 ; E - the modulus of elasticity in N/mm^2 [5].

The difference between the deformations, the elasticity modules of the masonry and the plaster coating is the reason for the shear strains in the "masonry - plaster coating" contact

zone (Figs. 3e, 4a) and stresses (τ) (Fig. 4b), which predetermine the development of a crack in the contact zone:

$$\tau = \frac{\Delta T_1 \cdot \alpha_1 - \Delta T_2 \cdot \alpha_2}{\left(\frac{1}{E_1} + \frac{1}{E_2}\right)}, \quad (13)$$

where: τ - shear stress from temperature deformations, kgf/cm^2 ; $\Delta T_1, \Delta T_2$ - the temperature difference at the time of installation and operation of the plaster coating and masonry, $^\circ\text{C}$; α_1, α_2 - coefficient of thermal expansion of masonry and plaster coating; E_1, E_2 - moduli of elasticity of masonry and plaster coating, kgf/cm^2 .

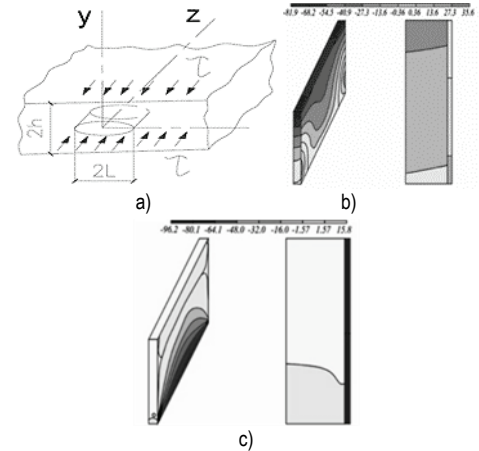


Figure 4 Deformations (a), stresses isopole from PC LIRA in the contact zone between the plaster coating and the masonry (b), (c)

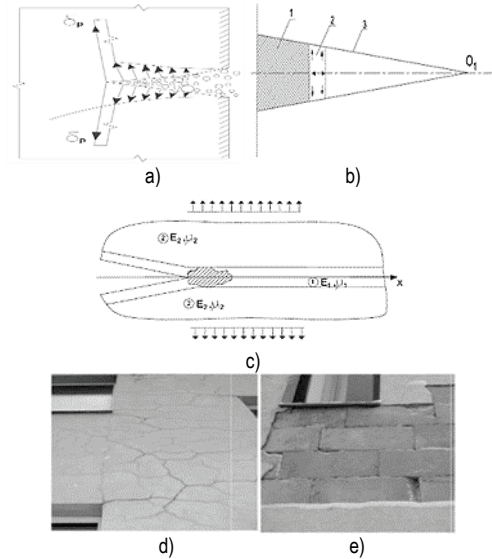


Figure 5 Development of the main crack in the contact zone due to the action of water (a), ice (b, c), and destruction of the plaster coating (d, e)

Atmospheric moisture, penetrating into the cracks of plaster coating, and through them into the contact zone, creates a wedging pressure at the top of the crack. Due to this, tensile stresses appear in this zone (Fig. 5a), which leads to a

further development of cracks in the plaster coating in its contact zone with the masonry.

At temperatures below zero degrees Celsius, the development of cracks is accelerated due to the transformation of water into ice (Figs. 5b, 5c), which ultimately leads to the destruction of the plaster coating and the wall structure (Fig. 5d, 5e).

3 EXPERIMENTAL STUDY OF PLASTER MORTAR

Reducing the elastic modulus of the stucco mortar and increasing its deformability are used in the case of fine aggregate with a low elastic modulus. In the experiment, to reduce the shrinkage of plaster coating and prevent the appearance of shrinkage cracks, a small filler and a filler with a low modulus of elasticity (vermiculite and from aerocrete fights (mix No. 1), limestone stone waste and perlite waste (No. 2 mix) were used.

The optimization of prescription and technological factors was carried out using a five-factor plan (Tab.1). For the experiment, the main factors were determined: X1 - Binder consumption (cement and lime mixture), X2 - Filler and aggregate consumption (perlite, carbonate and their mixture), X3 - Fiber consumption (diameter 18,7 μm , length 12 mm), X4 - redispersible polymer powder consumption Winnapas 5043 H, X5 - Consumption Tylose MBZ 15009. Their presence will also enable the reduction of the modulus of the elasticity of the plaster coating and the stress in it and the contact zone, and with the addition of a polymer fiber, the rate of development of cracks will be reduced, which will increase the durability of the coating and the wall structure.

The crack resistance of plaster mortars was determined in accordance with the National Standards of Ukraine (DSTU B B.2.7-126: 2011), while applying the mortar mixture to the aerated concrete foundation by using a frame device. A positive evaluation of crack resistance is no crack on the samples after curing for 24 hours under air-dry conditions. Since the crack resistance in the plaster mortar according to the National Standards of Ukraine (DSTU B B.2.7-126: 2011) are qualitative in nature, it was calculated that it is necessary to quantify the crack resistance coefficient (C_{cr}) of the formula:

$$C_{cr} = \frac{R_b}{R_c}, \quad (14)$$

where: R_b - bending strength; R_c - compressive strength plaster.

To determine the crack resistance of the plaster coating, its deformations, crack formation, and fracture were simulated on expanding deformable masonry (expansion deformations). For this purpose, cylinder samples were formed from the studied compositions (outer diameter 240 mm, inner diameter 200 mm, wall thickness 20 mm, height 150 mm). Tensile deformations in plaster coating were made by using two conical installations (Fig. 6a-6c).

To determine the properties of plaster mortar, more than 500 samples of a 40×40×160 mm beam were made on a gas-concrete base.

To simulate the real conditions, plaster coating was left to air-dry, and after 28 days of hardening, compressive strength, bending strength and density were determined (Tab. 2). Fracture toughness was determined visually by the presence of cracks in the coating and by calculating the fracture toughness ratio as the ratio of the bending strength in relation to compressive strength.

The study used an experimental statistical model (ESM) developed by V. Voznesensky and T. Lyashenko [11]. Based on the data obtained in the experiment, adequate experimental statistical (ES) models were constructed which describe the influence of variable factors on the properties of stucco mortars. The calculation of the model and its geometric display were carried out in the COMPLEX system. The result is plaster mortar with the following properties: Composition No. 1: average density 600-1500 kg/m^3 , bending strength 12-18 kg/cm^2 , compressive strength 18-36 kg/cm^2 , crack resistance coefficient 0,56-0,74 (Fig. 7a-d).

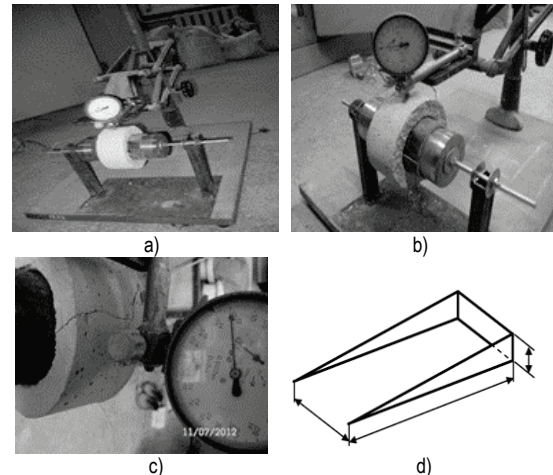


Figure 6 Installation for determining crack resistance (a-c), the framework for determining the crack resistance of the solution (d)

Table 1 Factor variation range

Type of mixture	The range of variation	Binder consumption, kg/m^3	Filler and aggregate consumption, kg/m^3	Fiber consumption, kg/m^3	Redispersible polymer powder consumption Winnapas 5043 H, %	Consumption Tylose MBZ 15009 %
		X1	X2	X3	X4	X5
Mix № 1	1	500	1,05/1	1,2	5	0,5
	0	400	1,05/1	0,9	3	0,3
	-1	300	1,05/1	0,6	1	0,1
Mix №2	1	400	1,05/1	1,2	5	0,5
	0	300	1,05/1	0,9	3	0,3
	-1	200	1,05/1	0,6	1	0,1

Table 2 Physical-mechanical properties of the plaster mortar

№	Mortar properties				
	Water retention %	R_b , kgf/cm ²	R_c , kgf/cm ²	W_p , kgf/cm ²	Average density, kg/m ³
1	94	8	10	0.68	555
2	97	16	18	0.74	711
3	97	10	12	0.58	1486
4	98	20	28	0.56	1512
5	91	9	14	0.72	490
6	93	16	21	0.78	723
7	99	15	20	0.4	1617
8	98	17	21	0.64	1496
9	98	13	14	0.76	510
10	96	20	27	0.24	703
11	97	24	38	0.38	1611
12	98	26	30	0.24	1568
13	95	14	16	0.66	570
14	94	22	24	0.68	678
15	97	25	44	0.58	1592
16	95	38	65	0.38	1619
17	97	17	20	0.48	1109
18	96	26	36	0.72	1166
19	98	15	20	0.74	623
20	94	21	34	0.4	1525
21	95	23	38	0.56	1284
22	95	19	22	0.6	1305
23	97	16	19	0.74	1072
24	97	20	27	0.62	1113
25	97	18	23	0.28	1111
26	99	15	20	0.58	1094
27	95	16	18	0.6	1078

Where: Water retention capacity - the water retention capacity of a mortar mixture, R_b - bending strength of hardened mortar, R_c - compressive strength of hardened mortar, W_p - water absorption by weight in 24 hours, kg/m² of hardened mortar.

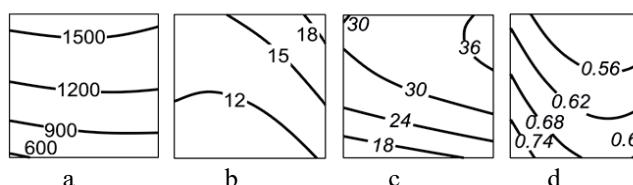


Figure 7 Isopole of basic physical and mechanical properties of the physical-mechanical properties of plaster mortar: average density (a), flexural strength (b) and compression (c), crack resistance (d) composition No. 1

Composition No. 2: average density 700-1100 kg/m³, bending strength 10-25 kg/cm², compressive strength 15-35 kg/cm², crack resistance coefficient 0,25-1 (Fig. 8a-8d).

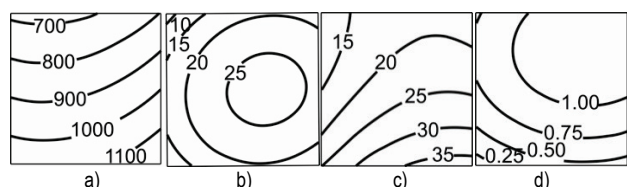


Figure 8 Isopole of basic physical and mechanical properties of the physical-mechanical properties of plaster mortar: average density (a), flexural strength (b), compressive strength (c), crack resistance coefficient (d) composition No. 2

The obtained plaster mortar meets the requirements of normative documents for the medium density (600-1600

kg/m³), flexural strength (0,98 -2,45 MPa) and compression (25-50 kg/cm²). The plaster coating has high crack resistance, the crack resistance coefficient is 0,25-1, while the cracking resistance is considered for the plaster with an index >0,26.

4 CONCLUSIONS

An increase in the share of high-rise buildings and the widespread use of new wall materials requires the development of scientific methods for the design of plaster mortar compositions. For this analysis, processes occurring in the plasters during their application and hardening were analyzed, failure mechanism for "bricklaying - plaster coating" systems were considered, and formula for calculating stress is given. The principles and criteria under which the durability of the plaster coating and the wall structure will be provided were formulated, the constituents were selected and the composition of the material was selected.

5 REFERENCES

- [1] Paruta, V. (2012). Theoretical premises for optimizing the formulation and technological parameters of plaster mortars for walls made of aerated concrete blocks. *Civil Engineering Journal*, 30-36. <https://doi.org/10.5862/MCE.34.4>
- [2] Galkin, S. (2006). *Application of cellular concrete products. Theory and practice*, NP Strinko LLC.
- [3] Knat'ko, M. (2008). On the Question of Durability and Energy Efficiency of Modern Enclosing Wall Structures of Residential, Administrative, and Industrial Buildings. *Engineering and Construction Journal*, 50-53.
- [4] Voldrich, F. (1978). *Deformation joints in the construction of above-ground buildings*, Stroyizdat.
- [5] Emelyanov, A. (1964). Damage to the exterior panels of residential prefabricated buildings with temperature deformations according to field studies. *Analysis of the causes of accidents and damage to building structures*, 153-177.
- [6] Blazi, V. (2005). *Reference Designer*. Construction Physics.
- [7] Rusova, R. (2001). Basis of wall and finishing materials production.
- [8] Sazhneva, N. (2009). Protective systems for the finishing of cellular concrete of low density. *Building materials*, 17-19.
- [9] Paplavskis, J. (2010). Requirements for plaster for exterior wall finishing of cellular concrete. Problems of the operational reliability of exterior walls based on autoclaved aerated concrete blocks and the possibility of their protection against moisture. *Plaster compositions for exterior decoration of aerated concrete walls: workshop materials*, 10-15.
- [10] Yemelyanov, A. (1964). Damage to the exterior panels of residential prefabricated buildings during thermal deformations according to field studies. *Analysis of the causes of accidents and damage to building structures*, 153-177.
- [11] Voznesensky, V. & Lyashenko, T. (1989). Numerical methods for solving construction and technological problems on a computer.

Authors' contacts:

Zeljko KOS, PhD, Ass. Prof.
University North, University Centre of Varaždin,
104. brigade 3, 42000 Varaždin, Croatia
+38598757989; zeljko.kos@unin.hr

Valentyn PARUTA, PhD, Ass. Prof.
Odessa State Academy of Civil Engineering and Architecture,
4, Didrikhson str., Odesa, Ukraine
+380676358991, docent2155@gmail.com

Olga GNYP, PhD, Ass. Prof.
Odessa State Academy of Civil Engineering and Architecture,
4, Didrikhson str., Odesa, Ukraine
+380679157473, asp@ogasa.org.ua

Leonid LAVRENYUK, PhD, Ass. Prof.
Odessa State Academy of Civil Engineering and Architecture,
4, Didrikhson str., Odesa, Ukraine
leonid-2319@ukr.net

Iryna GRINYOVA, PhD
Odessa State Academy of Civil Engineering and Architecture,
4, Didrikhson str., Odesa, Ukraine
+380939799301, irene.grinyova@gmail.com

DETECTION AND CLASSIFICATION OF BRAIN TUMOURS FROM MRI IMAGES USING FASTER R-CNN

Ercan AVŞAR, Kerem SALÇIN

Abstract: Magnetic resonance imaging (MRI) is a useful method for diagnosis of tumours in human brain. In this work, MRI images have been analysed to detect the regions containing tumour and classify these regions into three different tumour categories: meningioma, glioma, and pituitary. Deep learning is a relatively recent and powerful method for image classification tasks. Therefore, faster Region-based Convolutional Neural Networks (faster R-CNN), a deep learning method, has been utilized and implemented via TensorFlow library in this study. A publicly available dataset containing 3,064 MRI brain images (708 meningioma, 1426 glioma, 930 pituitary) of 233 patients has been used for training and testing of the classifier. It has been shown that faster R-CNN method can yield an accuracy of 91.66% which is higher than the related work using the same dataset.

Keywords: Brain Tumour; Classification; Convolutional Neural Network; Deep Learning; Glioma; Meningioma; Pituitary

1 INTRODUCTION

Cancer is one of the major causes of death today. According to the reports of World Health Organization (WHO), it is estimated that 9.6 million people worldwide died of cancer in 2018 (<https://www.who.int/cancer/en/>). 30-50% of these were preventable with early diagnosis. Between types of cancer, brain tumour is one of the deadliest ones. According to statistics, it is estimated that 17,760 adults will die from brain tumours in 2019 [1].

The complex structure of the human brain complicates the diagnosis of the tumour in the brain region. MRI, a useful method for obtaining high quality brain images, is widely utilized for tumour diagnosis. Especially in brain imaging, the MRI method provides a unique appearance in visualization of soft tissues with spatial resolution and contrast resolution.

Detection of brain tumours may be considered as an image segmentation problem where the tumour is labelled on the image. In order to solve this problem, various image processing methods as well as machine learning algorithms have been applied to MRI images by the researchers. Among the studies based on image processing, transform-based approaches and thresholding methods are common. For example, Salem and Alfonso utilized Fast Fourier Transform (FFT) and Minimal Redundancy-Maximal Relevance (MRMR) techniques are used for automatic classification of MRI brain tumour images [4]. In another work, Remya et al. used discrete wavelet transform (DWT) together with Fuzzy C-Means method for segmentation of tumours from MRI images [5]. While thresholding and clustering are common methods for tumour segmentation [6-8], it is also possible to combine these with transform-based methods [9].

Even though it is possible to detect tumours using these methods, it takes more processing to classify the tumours. For this purpose, methods based on machine learning are utilized. Shree and Kumar used Berkeley Wavelet Transformation (BWT) and Support Vector Machine (SVM) to detect normal and abnormal tissues from MRI brain

images [10]. In the study by Kumar and Vijayakumar the principle component analysis (PCA) and Radial Basis Function (RBF) are used for segmentation and classification of brain tumour and obtained better results [11]. Another well-known method for medical image classification is artificial neural networks (ANN). Ullah et al., used ANN to classify MRI brain images as normal or abnormal. The features of the images were extracted by Haar wavelet and statistical moments [12]. For a similar purpose, Muneer and Joseph utilized ANN together with (PCA) to reduce the dimensionality of the feature vector [13].

Deep learning is one of the machine learning methods that uses neural network architecture with possibly hundreds of hidden layers between the input and output layers [14]. It has been applied in various problems such as image classification [15], object detection [16, 17] and speech recognition [18, 19]. A common deep learning architecture is convolutional neural networks (CNN) where one of three types of operations are performed: convolution, pooling, or rectified linear unit (ReLU). A typical CNN can decide whether an image contains an object but without its location information. On the other hand, region-based CNN (R-CNN), an extended version of CNN, is mainly used for locating objects in images [20].

Deep learning is a highly effective method in solving medical image analysis problems including lung cancer diagnosis [21], tibial cartilage segmentation [22] and brain tumour detection [23-26]. Using deep learning methods, automatic segmentation has been successfully performed on large amounts of MRI images [23-25]. In particular, CNN-based algorithms for automated MRI segmentation for brain tumour yielded successful results by distinguishing distinctive features [26]. Among more recent solutions to this problem, Sajid et al. proposed a CNN architecture that takes local and contextual information into account [27]. Their method involves a pre-processing step to normalize the images and a post-processing step to eliminate the false positives. The obtained sensitivity and specificity values for glioma detection are 0.86 and 0.91, respectively. In another

study for brain tumour detection, a hybrid method utilizing Neutrosophy together with CNN is proposed [28]. The performance of the method is compared with regular CNN, SVM and K-nearest neighbours (KNN) and reported to be outperforming them with an accuracy of 95.62%.

In this work, MRI brain images are analysed using faster R-CNN method to detect and locate tumours in the images. Also, the detected tumours are classified into one of the tumour categories: meningioma, glioma, and pituitary. This method has been chosen because it can perform classification with higher accuracy and speed than regular R-CNN. Also, performance of faster R-CNN method for detecting the type of the tumour has not been investigated on this problem yet. Besides, effects of selecting different thresholds at the output layer have been analysed in detail by computing several performance metrics such as accuracy, precision, sensitivity, specificity, and f-score.

All the coding is performed with Python programming language (version 3.6.6) and implementation of deep learning model is done via TensorFlow library (<https://www.tensorflow.org/>).

The rest of the paper is organized as follows. Theoretical background information about faster R-CNN is mentioned in Section 2. Section 3 contains explanation of the dataset used as well as details of model training. Next, results and discussion are given in Section 4 and the paper is concluded with Section 5.

2 BACKGROUND

The process of building a CNN involves six main steps: Convolution, Rectified Linear Unit (ReLU), Pooling, Flattening, Fully connected layers, and Softmax function (Fig. 1). At the initial step, a number of convolution filters are applied to the input image to activate certain features from the images. In order to increase the training speed, negative values are mapped to zero and positive values remained unchanged in ReLU step. The purpose of pooling step is to simplify the output by performing nonlinear downsampling, and hence reducing the number of parameters that the network needs to learn about. These three operations are repeated over tens or hundreds of layers, with each layer learning to detect different features. In the flattening step, all two-dimensional arrays are transformed into one single linear vector. Such a process is needed for fully connected layers to be used after convolutional layers. Fully connected layers are able to combine the entire local features of the previous convolutional layers. The procedure is finished with application of softmax function to provide the final classification output.

It is possible to tell the class of the objects within an image via CNN; however, it does not return any location information. In order to find the object location R-CNN method has been developed. In R-CNN, some candidate regions of the image are selected with some region proposal functions. These regions are cropped and resized to be classified by CNN method. One improved method of R-CNN is fast R-CNN where the entire image is processed rather than cropping the image and resizing the cropped parts. Region

proposal functions used in fast R-CNN as well. However, unlike CNN, these regions are not processed directly; CNN features corresponding to each region proposal are pooled by fast R-CNN method. Generation of region proposals in a faster and better way is achieved by region proposal network (RPN). Faster R-CNN method introduces RPN directly in the network thereby enables a higher classifier performance [29].

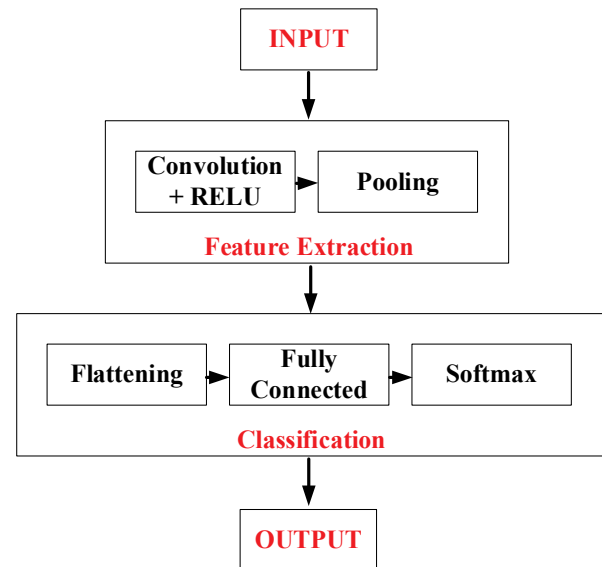


Figure 1 Structure of CNN

3 MATERIALS AND METHOD

3.1 The Dataset

The brain image dataset contains 3,064 MRI slices from 233 patients. Each of these images contains one of the three types of brain tumours: meningioma, glioma, or pituitary. The images have an in-plane resolution of 512×512 with pixel size 0.49×0.49 mm². The slice thickness is 6 mm and the slice gap is 1 mm [11]. Sample images from each of the classes are shown in Fig. 2 and the distribution of the images having these tumour types is given in Tab. 1.

Table 1 Distribution of the dataset in three classes

Tumour Type	Number of Images
Meningioma	708
Glioma	1426
Pituitary	930
Total:	3064

3.2 Training

In order to train and test the faster R-CNN classifier, the tumour locations in all of the images should be labelled and saved to a file. For this purpose, graphical image annotation tool called Labellmg [30] is used. Labelimg is written in Python and uses Qt for its graphical interface. Annotations are saved as XML files in PASCAL VOC format, the format used by ImageNet (<http://www.image-net.org/>).

After labelling process, 80% of the images in the database are randomly selected as training samples. In order to prevent the training set to be dominated by one of the

classes, same proportion of images is selected from each class. The remaining 20% is left for testing. A faster R-CNN model is trained on the training data. The parameters regarding the model are as follows:

- Learning rate: 0.0002
- Momentum constant: 0.9
- Batch size: 1
- Number of iterations: 150,000
- Softmax neurons: 3
- Max-pool kernel size: 2
- Max-pool stride: 2

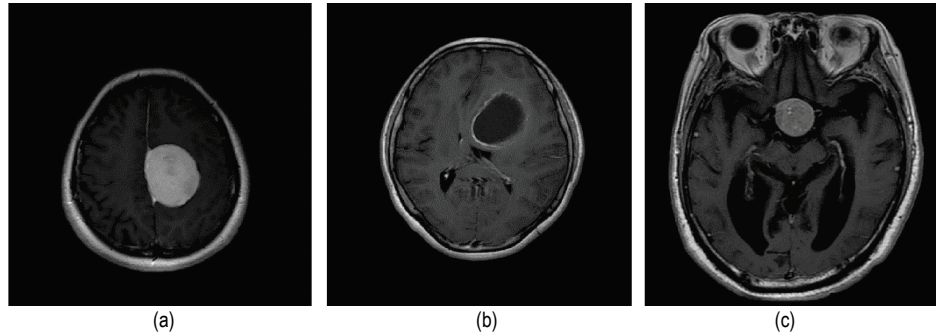


Figure 2 Images containing different types of tumours: (a) Meningioma, (b) Glioma, (c) Pituitary

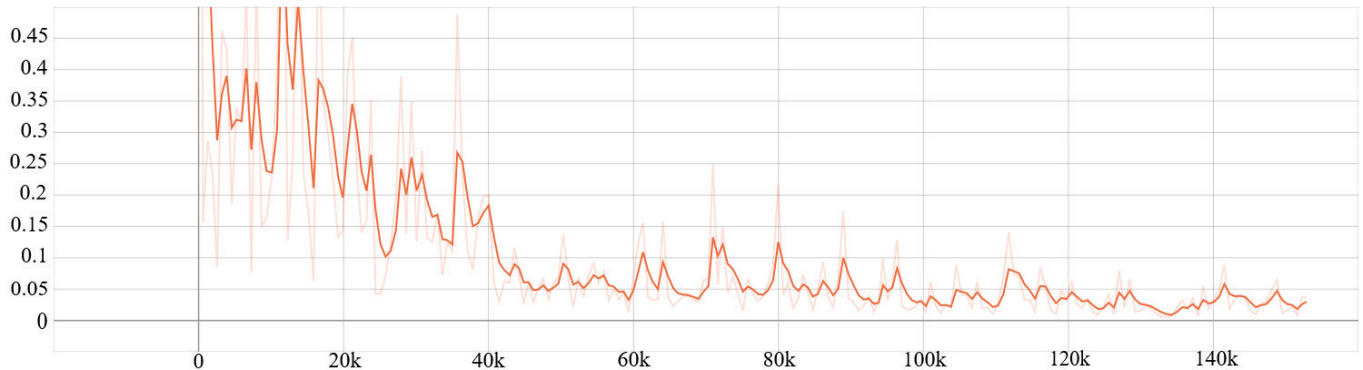


Figure 3 Total Loss Graph of the system (Vertical axis: total loss, Horizontal axis: number of iterations)

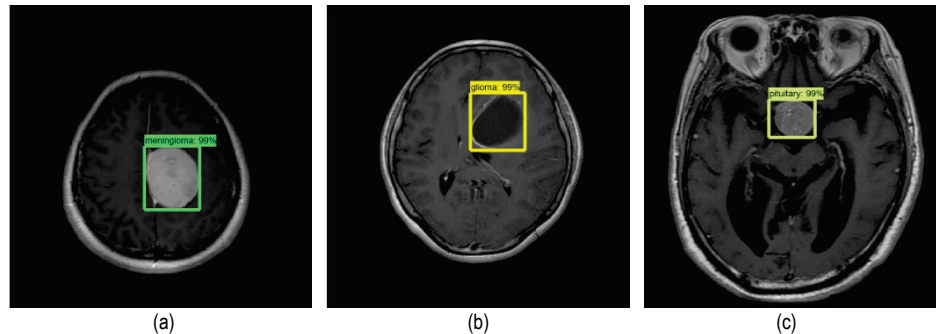


Figure 4 Detected and classified brain tumours: (a) Meningioma, (b) Glioma, (c) Pituitary

The loss function that is being minimized during training is provided in Fig. 3. At the beginning of the training, very high loss rate is observed. For example, in the first iteration total loss is approximately 90%. After 140.000 iterations, total loss is decreased to a value about 2%. This is an expected situation indicating that the algorithm performs learning process. Basically, two different loss functions are minimized in this problem. One is related to the correct classification rate and the other one is related to the location of the correctly classified object. Since graphs of both functions are similar, only Total Loss graph is given here as the sum of these two functions.

4 RESULTS AND DISCUSSION

The 612 samples (constituting 20% of the dataset) are used for testing the faster R-CNN model. For each test image three parameters are returned: (i) class label of the detected tumour, (ii) probability of the tumour being in that class and (iii) location of the tumour. Sample classification results are shown in Fig. 4.

In TensorFlow library, the minimum softmax probability value for assigning a class label to a detected tumour is set as 0.8 by default. This means that a tumour is left unclassified if the associated probability value is smaller than 0.8. By changing this probability threshold, we observed the change

in number of unclassified and misclassified samples. Three confusion matrices belonging to three different threshold values (0.8, 0.66, and 0.5) are given in Tabs. 2, 3, and 4, respectively.

Five performance metrics, given in equations 1-5, are calculated using the classification results. Computation of these quantities requires definition of the confusion matrix elements, that are True Positive (*TP*), True Negative (*TN*), False Positive (*FP*) and False Negative (*FN*). Since values of these elements change according to the reference class label, three different values for each of the performance metrics are calculated for each of the confusion matrices in Tabs. 2, 3, and 4. The related results are given in Tabs. 5, 6, and 7.

Table 2 Confusion Matrix when threshold is 0.8 (63 samples are unclassified)

PREDICTED	ACTUAL			
		Meningioma	Glioma	Pituitary
	Meningioma	129	5	2
	Glioma	0	237	1
	Pituitary	4	2	169
UNCLASSIFIED		8	41	14

Table 3 Confusion Matrix when threshold is 0.66 (48 samples are unclassified)

PREDICTED	ACTUAL			
		Meningioma	Glioma	Pituitary
	Meningioma	130	6	3
	Glioma	0	244	1
	Pituitary	6	2	172
UNCLASSIFIED		5	33	10

Table 4 Confusion Matrix when threshold is 0.5 (35 samples are unclassified)

PREDICTED	ACTUAL			
		Meningioma	Glioma	Pituitary
	Meningioma	133	6	3
	Glioma	0	251	1
	Pituitary	4	2	177
UNCLASSIFIED		4	26	5

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

$$Precision = TP/(TP + FP) \quad (2)$$

$$Sensitivity = TP/(TP + FN) \quad (3)$$

$$Specificity = TN/(FP + TN) \quad (4)$$

$$F-score = 2*TP/(2*TP + FP + FN) \quad (5)$$

Table 5 Performance metrics when threshold is 0.8

Reference Class	Meningioma	Glioma	Pituitary
Accuracy	0.8742	0.8742	0.8742
Precision	0.9485	0.9958	0.9086
Sensitivity	0.9149	0.8316	0.9657
Specificity	0.9831	0.9967	0.9556
F-Score	0.9314	0.9063	0.9363

Table 6 Performance metrics when threshold is 0.66

Reference Class	Meningioma	Glioma	Pituitary
Accuracy	0.8922	0.8922	0.8922
Precision	0.9353	0.9959	0.9570
Sensitivity	0.9220	0.8561	0.9271
Specificity	0.9788	0.9967	0.9796
F-Score	0.9286	0.9208	0.9418

Table 7 Performance metrics when threshold is 0.5

Reference Class	Meningioma	Glioma	Pituitary
Accuracy	0.9166	0.9166	0.9166
Precision	0.9500	0.9960	0.9674
Sensitivity	0.9433	0.8811	0.9519
Specificity	0.9840	0.9968	0.9847
F-Score	0.9466	0.9351	0.9596

It is clearly seen from the confusion matrices that the number of unclassified samples decreases as the decision threshold is decreased. Also, the related performance measures are improved by decreasing the threshold. Considering the accuracy values in Tabs. 5, 6 and 7, 87.42% of the images may correctly be classified when the decision threshold is set to 0.8 and this accuracy is increased to 91.66% when the threshold is set as 0.5. This means that, even though the output probability of the classifier is relatively low for some of the images, the tumour can be correctly detected. On the other hand, majority of the unclassified and misclassified samples belong to "Glioma" class, meaning that the classifier cannot learn this class as effectively as the other two. However, specificity value related to "Glioma" class is the highest for three of the thresholds.

5 CONCLUSION

In this study, MRI brain images have been used for detection of tumours using a deep learning method, namely faster R-CNN. A model has been trained using 2452 images containing three types of tumours. It may be concluded that faster R-CNN is suitable algorithm for this problem as the achieved classification accuracy is 91.66% which is higher than the related work using the same dataset. Also, it has been observed that some of the tumours are correctly detected with low scores. Therefore, the results are evaluated using three different levels of detection scores. Among the three types of tumours, Glioma has the lowest detection rate with a sensitivity of 88.11% and Pituitary has the highest rate with a sensitivity of 95.19%. Since only one test sample is predicted as false positive for Glioma class, the precision value of the Glioma is the highest, on the other hand high number of false negative samples causes its sensitivity value to be low.

One major advantage of deep learning is that the classification accuracy is improved as the number of training examples is increased. However, this is not the case for earlier machine learning methods in which no increment is observed above a certain amount of training examples. Among deep learning methods, faster R-CNN has been proven to be performing faster with a higher classification accuracy [29]. This method also returns the class label and location of the object together. These properties of faster R-CNN make it suitable to be applied to tumour detection problem where it is important to show the location of the tumour. Therefore, contributions of the study may be listed as: (i) There is a potential for improving the performance of the method when more data is available, (ii) tumours are located and classified simultaneously, (iii) classification

performance is improved by changing the threshold value at the output layer.

A careful attention should be paid to the unclassified and misclassified samples. The presence of unclassified samples is something related to detections with low scores. Such samples may be thought to be classified as “no tumours” even though they contain a tumour. In the future, this situation may be overcome by adding healthy images (i.e. images containing no tumours) to the dataset as the fourth class. As for reduction of the number of misclassified samples (particularly the samples from Glioma class), a pre-processing step is planned to be added to the proposed method. The pre-processing step may include algorithms that will emphasize the unobvious features in the given images. Unsharp masking and histogram equalization are two possible examples of these algorithms. Besides, as can be seen from Section 3.2, a set of constant parameters has been used for generating the CNN model. In the future, some optimization algorithms (such as genetic algorithm, particle swarm optimization, simulated annealing, etc.) may be utilized to find the best parameter set giving the highest classification accuracy.

Acknowledgement

This work was supported by the Research fund of the Çukurova University under grant no FYL-2019-11415.

6 REFERENCES

- [1] Cancer.Net Editorial Board. (2019). Brain Tumor: Statistics. Retrieved from <https://www.cancer.net/cancer-types/brain-tumor/statistics>
- [2] Veer, S. & Patil, P. (2015) Brain Tumor Classification Using Artificial Neural Network on MRI Images, *International Journal of Research in Engineering and Technology*, 4(12), 218-226. <https://doi.org/10.15623/ijret.2015.0412042>
- [3] Demirhan, A. & Güler, I. (2014). Using Wavelet Transform and Neural Networks for the Analysis of Brain MR Images. *SIU2010 - IEEE 18. Sinyal isleme ve iletisim uygulamaları kurultayı*. 933-936. <https://doi.org/10.1109/SIU.2010.5651477>
- [4] Alfonse, M. & Salem, A.-B. M. (2016). An automatic classification of brain tumors through MRI using support vector machine. *Egyptian Computer Science Journal*, 40(3), 11-21.
- [5] Remya, R., Parimala, G. K., & Sundaravadivelu, S. (2019). Enhanced DWT Filtering Technique for Brain Tumor Detection. *IETE Journal of Research*, 1-10. <https://doi.org/10.1080/03772063.2019.1656555>
- [6] Wu, M., Lin, C., & Chang, C. (2007). Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation. *Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007)*, 2, 245-250. <https://doi.org/10.1109/IIH-MSP.2007.4457697>
- [7] Natarajan, P., Krishnan, N., Kenkre, N. S., Nancy, S., & Singh, B. P. (2012). Tumor detection using threshold operation in MRI brain images. *IEEE International Conference on Computational Intelligence and Computing Research*, 1-4. <https://doi.org/10.1109/ICCIC.2012.6510299>
- [8] Mustaqeem, A., Javed, A., & Fatima, T. (2012). An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation. *International Journal of Image, Graphics and Signal Processing*, 4(10), 34-39. <https://doi.org/10.5815/ijigsp.2012.10.05>
- [9] Deepa, B. & Sumithra, M. G. (2019). An intensity factorized thresholding based segmentation technique with gradient discrete wavelet fusion for diagnosing stroke and tumor in brain MRI. *Multidimensional Systems and Signal Processing*, 30(4), 2081-2112. <https://doi.org/10.1007/s11045-019-00642-x>
- [10] Shree, N. V. & Kumar, T. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. *Brain Informatics*, 5(1), 23-30. <https://doi.org/10.1007/s40708-017-0075-5>
- [11] Kumar, P. & Vijayakumar, B. (2015). Brain Tumour MR Image Segmentation and Classification Using by PCA and RBF Kernel Based Support Vector Machine. *Middle-East Journal of Scientific Research*, 23(9), 2106-2116.
- [12] Ullah, Z., Lee, S. H., & Fayaz, M. (2019). Enhanced feature extraction technique for brain MRI classification based on Haar wavelet and statistical moments. *International Journal of Advanced and Applied Sciences*, 6(7), 89-98. <https://doi.org/10.21833/ijaas.2019.07.012>
- [13] Muneer, K. V. A. & Joseph, K. P. (2019). Automation of MR Brain Image Classification for Malignancy Detection. *Journal of Mechanics in Medicine and Biology*, 19(1), Art. no. 1940002 <https://doi.org/10.1142/S0219519419400025>
- [14] Şeker, S., Diri, B., & Balık, H. (2017). Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme (In Turkish). *Gazi Mühendislik Bilimleri Dergisi*, 3(3), 47-64.
- [15] Chan, T., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A Simple Deep Learning Baseline for Image Classification. *IEEE Transactions on Image Processing*, 24(12), 5017-5032. <https://doi.org/10.1109/TIP.2015.2475625>
- [16] Ouyang, W. et al. (2015). Deepid-net: Deformable deep convolutional neural networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, 2403-2412. <https://doi.org/10.1109/CVPR.2015.7298854>
- [17] Han, J., Zhang, D., Cheng, G., Liu, N., & Xu, D. (2018). Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey. *IEEE Signal Processing Magazine*, 35(1), 84-100. <https://doi.org/10.1109/MSP.2017.2749125>
- [18] Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 8599-8603. <https://doi.org/10.1109/ICASSP.2013.6639344>
- [19] Deng, L. & Platt, J. C. (2014). Ensemble deep learning for speech recognition. *15th Annual Conference of the International Speech Communication Association*.
- [20] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [21] Sun, W., Zheng, B., & Qian, W. (2017). Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Computers in Biology and Medicine*, 89, 530-539. <https://doi.org/10.1016/j.compbiomed.2017.04.006>
- [22] Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., & Nielsen, M. (2013). *Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural*

- Network*. International Conference on Medical Image Computing and Computer-Assisted Intervention, 246-253.
<https://doi.org/10.1007/978-3-642-40763-5>
- [23] Havaei M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., & Larochelle, H. (2016). Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35, 18-31.
<https://doi.org/10.1016/j.media.2016.05.004>
- [24] Işın, A., Direkçioğlu, C., & Şah, M. (2016) Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods. *Procedia Computer Science*, 102, 317-324.
<https://doi.org/10.1016/j.procs.2016.09.407>
- [25] Justin, S. P., Andre J. P., Bennett, A. L., & Fabbri, D. (2017). Deep learning for brain tumor classification. *Proc. SPIE Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 10137.
<https://doi.org/10.1117/12.2254195>
- [26] Urban, G., Bendszus, M., Hamprecht, F. A., & Kleesiek, J. (2014). Multi-modal Brain Tumor Segmentation using Deep Convolutional Neural Networks. *MICCAI BraTS (Brain Tumor Segmentation) Challenge*. Proceedings, winning contribution, 31-35.
- [27] Sajid, S., Hussain, S., & Sarwar, A. (2019). Brain Tumor Detection and Segmentation in MR Images Using Deep Learning. *Arabian Journal for Science and Engineering*, 44(11), 9249-9261. <https://doi.org/10.1007/s13369-019-03967-8>
- [28] Özyurt, F., Sert, E., Avci, E., & Dogantekin, E. (2019). Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. *Measurement*, 147, p. 106830.
<https://doi.org/10.1016/j.measurement.2019.07.058>
- [29] Ezhilarasi, R. & Varalakshmi, P. (2018). *Tumor Detection in the Brain using Faster R-CNN*. Proceedings of the Second International conference on I-SMAC.
<https://doi.org/10.1109/I-SMAC.2018.8653705>
- [30] Lin, T. (2015). Labelimg. <https://github.com/tzutalin/labelimg>

Authors' contacts:

Ercan AVŞAR, PhD
 Corresponding author
 Çukurova University,
 Department of Electrical and Electronics Engineering,
 Balcali, Saricam. Adana/Turkey
 +903223386868, ercanavsar@cu.edu.tr

Kerem SALÇIN, Graduate Student
 Çukurova University,
 Department of Electrical and Electronics Engineering,
 Balcali, Saricam. Adana/Turkey
 +903223386868, k.salcin@gmail.com

ANALYSIS OF THERMOMECHANICAL PROPERTIES OF POLYMERIC MATERIALS PRODUCED BY A 3D PRINTING METHOD

Adam GNATOWSKI, Rafał GOŁĘBSKI, Piotr SIKORA

Abstract: A comparative analysis of the thermomechanical properties of semicrystalline and amorphous polymeric materials was carried out. Samples were produced by using a 3D printing technology on the SIGNAL printer - ATMAT. The following polymeric materials were used to make the samples: TPU-thermoplastic polyurethane elastomer, ABS-copolymer acrylonitrile-butadiene-styrene, Nosewood, PET-ethylene terephthalate, PLA-poly (lactic acid). The research included a thermal analysis of the dynamic properties (DMTA) of manufactured materials.

Keywords: polymer materials; thermal analysis of dynamic properties; 3D print

1 INTRODUCTION

3D printed parts are more and more often used in different industry branches as specially prototyping and commonly used items. Those sectors use the printing technology due to the short time between the project and final item production, and also because of the lower prices of manufacturing small batches compared to the injection moulding technology. Furthermore, 3D printing gives an opportunity to change the shape of the next part only by adding some changes in the 3D model. It also gives us the possibility of creating very complicated shapes [1, 2]. Modern printers, due to the new software and hardware could spread subsequent filament layers more precisely, as they help them reach narrow dimensional and position tolerances of the printed elements. Among the tested materials used for 3D printing, you can find Laywood wood-like material, which after printing gives the elements the appearance of fibreboard and could be further processed by turning, milling grinding and could be painted as normal wood. This material is used to manufacture architectural models, decorations and furniture elements [3]. The amorphous acrylonitrile butadiene styrene used in the 3D printing technology is a high strength material since it works well when printing moving parts due to good temperature resistance and increased elasticity [4]. On the other hand, when this material is heated, it releases intense chemical vapours which are harmful for human being and animals, which is why such printing material should be placed in enclosed printers placed in a well-ventilated room. PLA is another test material, one of the most commonly used 3D printing materials characterized by biodegradability, since it is made of easily degradable materials. This material is characterized by a slight shrinkage of the material, which makes it ideal for demonstration prints such as mock-ups or prototype elements. PLA is not suitable for printing moving parts due to its low flexibility and low temperature resistance [5].

An important part of 3D printing is the appropriate choice of a material which will fulfil all our needs associated with our printed element. An incorrect selection of the material and process parameters associated with the type of element that will be manufactured and its functions may

cause problems, such as, for example, early wear of the element, difficulties in obtaining adequate model accuracy, inaccurate filling of internal spaces, etc. Another difficulty in the process of 3D printing is the selection of adequate parameters for filament type processes, such as the printing temperature or the working table temperature. Incorrect selection of these factors may result in: clogging printer nozzles, larger material shrinkage than the expected, incomplete plasticization when applying subsequent layers, resulting in delamination of the element and the occurrence of non-uniformity of the model filling which causes a decrease in its strength [6-8].

The aim of the paper is to analyse changes in the thermomechanical properties as a function of the temperature and vibration frequency of materials used to manufacture products by using 3D printing. The polymers were chosen because they are widely used in 3D printing. Some of these materials are not suitable for injection moulding due to the clogging of injection channels or too long exposure to higher temperatures, which changes the properties of the injected material such as Laywood. Tests were conducted for comparison purposes for significantly different materials. The following materials were used as a test material: ABS, Laywood, PLA, PET, TPU. A thermal analysis of the dynamic properties (DMTA) was performed. Researches show that thermomechanical properties of these materials are strictly dependent on the processing method. To obtain those comparisons, the paper examines the samples made by the injection moulding and printing technology [9-15]. In order to compare the properties of the samples produced by 3D printing and injection, samples were made from the ABS material, which is widely used in many industry branches. This enabled the determining of the differences between the properties of the tested materials obtained in various processing technologies.

2 RESEARCH METHODOLOGY

The DEVIL DESIGN filaments with a thickness of 1.75 mm were used in the research.

They were made of the following polymer materials: Laywood, ABS, PLA, PET, TPU. Samples were made by

using the 3D printing technology on an ATMAT SIGNAL printer, with the option of mounting nozzles with a through-going from 0.1 mm to 1.0 mm. A printer nozzle can warm up to 270 °C, while the table can reach a temperature of 110 °C. The accuracy of filament printing is 0.1 mm for the speed from 0.1 to 1 cm³/min. 3D printing was done by using a 0.8 mm nozzle and the DIMAFIX adhesive for printing was applied to the table and printing parameters were used, and the most favorable properties were obtained. Parameters at which 3D printing samples were produced guaranteed the stability of the printing process, no deformation and favourable thermomechanical properties (Tab. 1). The filament feeding speed was used in accordance with the manufacturer's recommendations, and print cooling was also applied. The layer height of 0.21 mm was adopted.

Table 1 Parameters of printing samples

Sample	Nozzle temperature, °C	Table temperature, °C	Printing speed, mm/s	Fulfilment, %
LAYWOOD	220	55	60	96
ABS	230	100	60	96
PLA	230	55	60	96
PET	242	60	50	96
TPU	250	60	25	96

Samples for comparative purposes made of injected ABS were produced on a Krauss Maffei type KM with a 30 mm diameter screw, LD ratio of 23, constant pitch over the entire length and mold closing force of 650 kN. The optimal properties of the specimens were obtained for the following injection parameters recommended by the material manufacturer:

- Injection temperature – 230 °C,
- Mold temperature – 60 °C,
- Holding pressure – 50 MPa,
- Maximum permissible pressure in the plasticizing system – 100 MPa,
- Time of clamping – 10 s,
- Cooling time – 16 s.

An analysis of the dynamic thermal properties was carried out with the NETZSCH DMTA 242 device with a handle for a three-point free bending sample in the form of a beam with the dimensions of 50×10×4 mm. The samples in the holder were subjected to a sinusoidal force of 1 Hz and 10 Hz with constant amplitude, while heating the samples at a rate of 3 °C/min from 20 °C to a different temperature for the materials tested (Tab. 2).

Table 2 Temperatures of the DMTA testing

Sample	DMTA
LAYWOOD	–50 °C - 60 °C
PLA	–50 °C - 60 °C
TPU	–50 °C - 170 °C
ABS	–50 °C - 190 °C
PET	–50 °C - 200 °C

The elastic modulus E' and the mechanical loss factor $\tan\delta$ were calculated on the basis of force and strain values in

relation to the dimensions of the samples. In order to be sure of the results, all tests were repeated for three samples.

3 DISCUSSION OF RESEARCH

The results of the analysis of dynamic mechanical properties are presented in Figs. 1 - 5a, 5b in the form of graphs of changes in the modulus of the elasticity and tangent of the angle of the mechanical losses depending on the temperature and the frequencies of 1 and 10 Hz. In the case of LAYWOOD, it was observed that the polymer is in the elastic deformation phase of about 54 °C. You can then see a quick transition to the glass transition phase state. At the beginning of the elastic deformation phase, the E' module is strongly dependent on the temperature of –30 °C and the relationship decreases significantly. In this phase, the E' module drops from 2700 MPa to 1800 MPa at 54 degrees Celsius. In this phase, it can be concluded that the tangent of the mechanical loss angle is characterized by a small dependence on the temperature, ranging from –10 °C to 54 °C. The transition to the elastoplastic state occurs rapidly at 54 °C, when the E' module rapidly decreases with a simultaneous sudden increase in $\tan\delta$ several times. This means that the polymer begins to bend and does not back to its normal shape, i.e. it bends permanently. That is why the machine could not read the data and that is why the chart ends at about 62 °C.

Fig. 2 is similar to Figure 1 because the LAYWOOD material is made of PLA with the addition of wood chips, which is why DMTA for PLA ends in the same way as LAYWOOD.

In the elastic deformation phase of PLA DMTA, which ends at –57 °C, the E' modulus shows a negligible temperature dependence and is shaped at 4800 MPa, while the $\tan\delta$ values increase almost twice, showing a strong temperature dependence. The glass transition phase starts at 51 °C and the chart ends at about 62 °C for reasons similar as to why LAYWOOD is characterized by an extension between –42 °C and 56 °C. In this phase, the $\tan\delta$ increases slightly to 0.08, and finally the phase is reduced to 0.05, followed by an increase in the tangent of the mechanical loss angle, which indicates the variability of vibration damping and material rigidity in the glass transition phase and a strong temperature dependence of the mechanical loss angle [9-12,16].

In the case of TPU (Fig. 3), the elastic deformation phase ends at –47 °C. The glass transition phase occurs in a wide temperature range, from –47 °C to 30 °C, in which E' and $\tan\delta$ show a strong temperature dependence. A high-viscoelastic deformation phase is extended in a wide temperature range: between 30 °C and 95 °C, and here you can see a strong temperature dependence of $\tan\delta$, which indicates the variability of vibration damping and material rigidity. The high-viscoelastic deformation phase begins with a rapid reduction in the value of $\tan\delta$ (from 0.17 to 0.04), which is associated with strong elasticity at ambient temperatures and a poor dependence of the E' module on the temperature. In the elastic transition phase, it starts at 95 °C, where the polymer begins to flow.

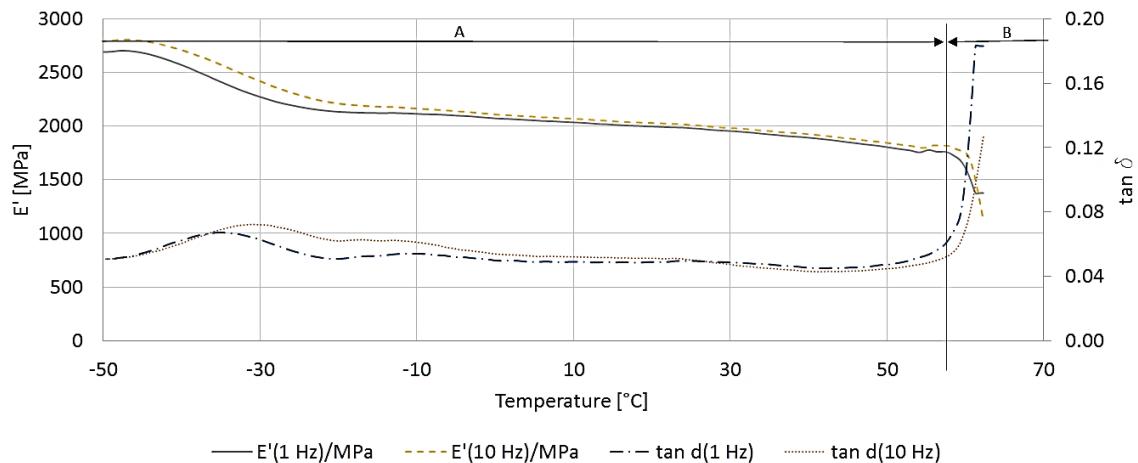


Figure 1 Results of DMTA investigations of LAYWOOD: A - elastic deformation phase; B - glass transition phase; C - high-viscoelastic deformation phase.

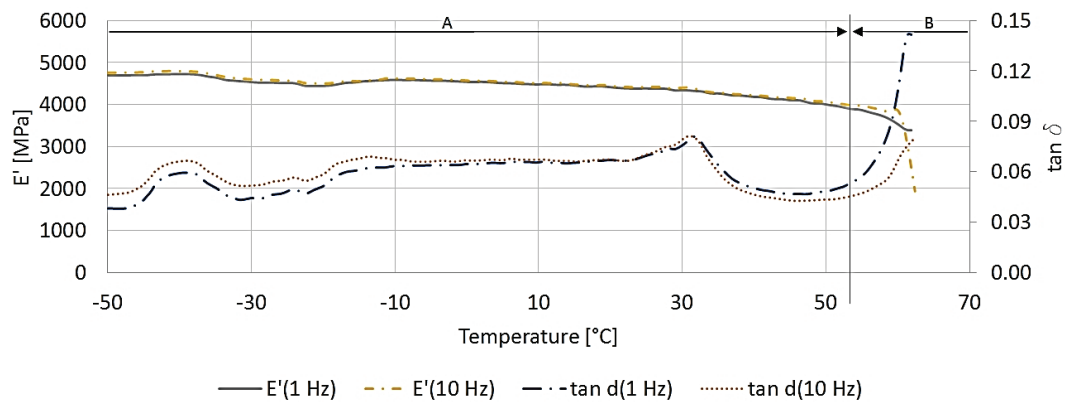


Figure 2 Results of DMTA investigations of PLA: A - elastic deformation phase; B - glass transition phase; C - high-viscoelastic deformation phase.

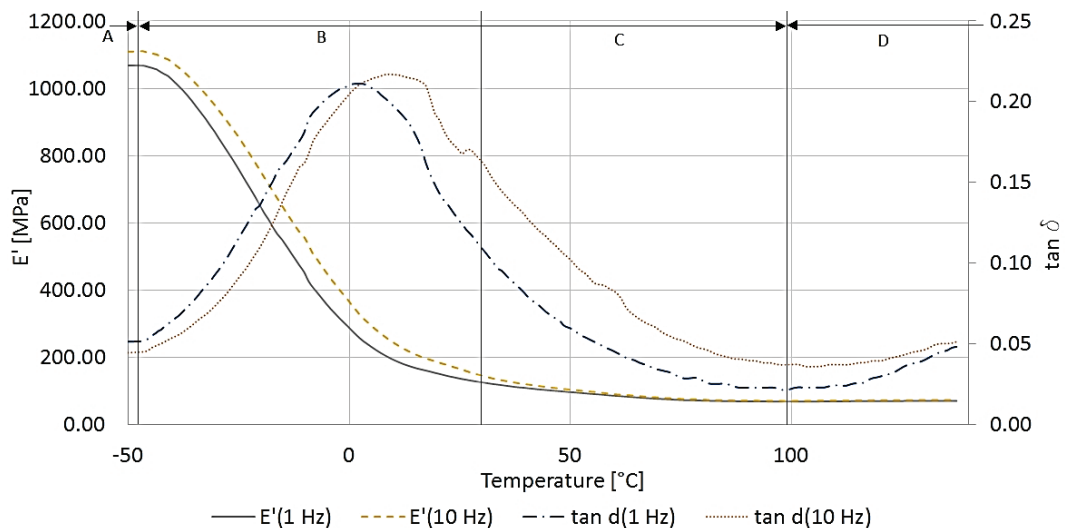


Figure 3 Results of DMTA investigations of TPU: A - elastic deformation phase; B - glass transition phase; C - high-viscoelastic deformation phase; D - elastic transition phase.

Fig. 4 shows the dependence of $\tan \delta$ and E' on temperature and frequency for PET. This material is a brittle and rigid body in the elastic deformation phase up to 73 °C, in which $\tan \delta$ and E' are poorly dependent on temperature. In the glass transition phase, the dependence of E' is substantial, because the E' modulus is reduced from 2400 MPa to 200 MPa, $\tan \delta$ shows a notable temperature dependence rising from 0.05 to 0.5 and at the end of the phase, it drops to 0.18

at the smallest frequency of extortion. The high-viscoelastic deformation phase begins with a rapid decrease of $\tan \delta$ (1 Hz) from 0.18 to 0.02 at 122 °C and at this temperature, the high-viscoelastic deformation phase begins [16-18]. Maximum $\tan \delta$ indicates that PET samples made in 3D printing do not show a high vibration damping ability and are characterized by low stiffness in relation to the other tested materials.

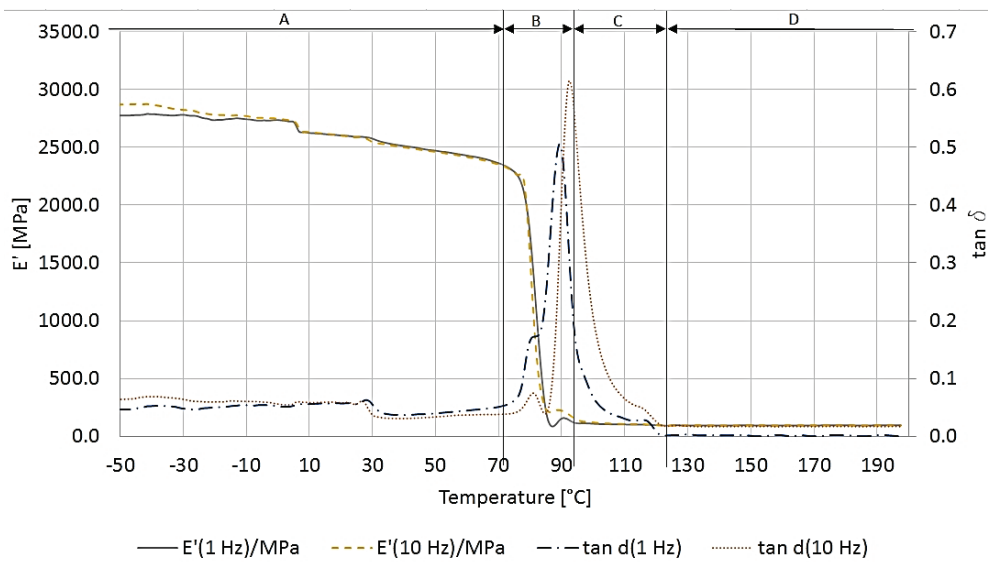


Figure 4 Results of DMTA investigations of PET: A - elastic deformation phase; B - glass transition phase; C - high-viscoelastic deformation phase; D - elastic transition phase.

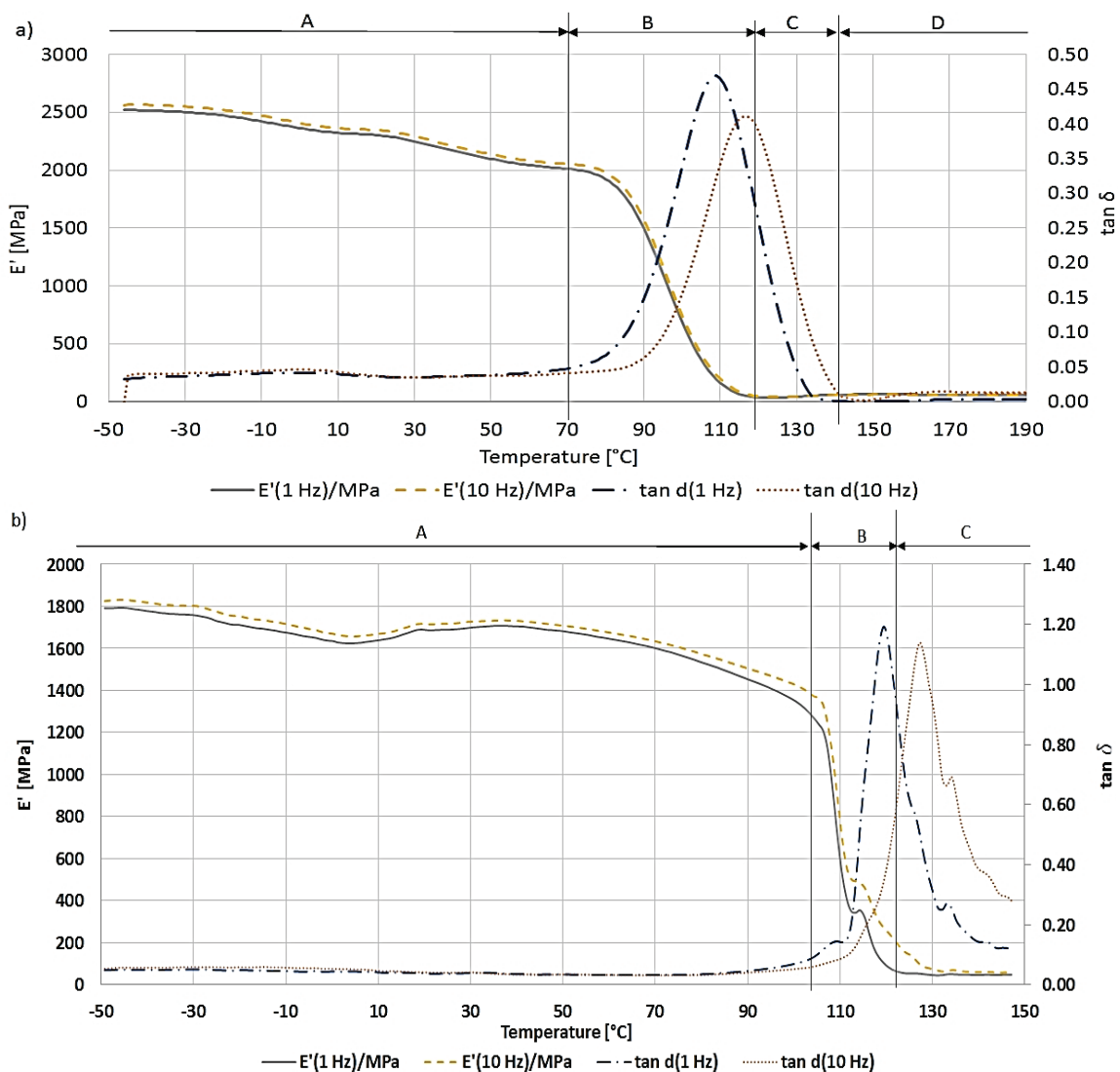


Figure 5 Results of DMTA investigations of: a) ABS from 3D printing, b) ABS from injecting: A - elastic deformation phase; B - glass transition phase; C - high-viscoelastic deformation phase.

The graph of Fig. 5a shows the course of changes of the module E' and $\tan\delta$ as a function of temperature and frequency of vibrations for ABS. In the case of ABS, the elastic deformation phase ends at 93 °C, smoothly going into the glass transition phase with a low temperature variation of E' and with a lack of $\tan\delta$ temperature dependence. The glass transition phase is characterized by almost tremendous temperature compliance at 93 °C, and the module E' drops from 1450 MPa to 100 MPa, $\tan\delta$ also shows temperature dependence because it goes from 0.1 to 1.2 and at the end of the phase drops to 0.6. In the high-viscoelastic deformation phase, $\tan\delta$ modules exhibit high temperature dependence, the $\tan\delta$ (1 Hz) drops to 0.1 at 146 °C. This polymer exhibits high stiffness and a very high vibration damping ability compared to the other tested samples as evidenced by the maximum $\tan\delta$ relaxation conversion for the smallest frequency of extortion.

The graph of Fig. 5b shows the course of changes of the module E' and $\tan\delta$ as a function of temperature and frequency of vibrations for the ABS samples obtained by injection molding. These samples have the A phase, which starts from 2500 MPa, and it is more than the ABS printed samples have; in phase B, the module E' drops very fast from 2000 MPa to 150 MPa, and this drop is faster than in the printing samples, but $\tan\delta$ in this phase is rising only to 0.45, i.e. it is 0.75 lower than in the 3D printed ABS. The phase C of the injected ABS is characterized by a fast drop of the $\tan\delta$ comparing to the printed ABS, because $\tan\delta$ reaches the lowest level at 134 °C, while printed ABS reaches the lower level of $\tan\delta$ at more than 150 °C [6, 11-14, 19]

4 CONCLUSIONS

An analysis of the results indicated that the samples made of Semi-Crystalline Polymers are characterized by an extended glass transition phase, in which the E' and $\tan\delta$ modules have strong temperature dependence. LAYWOOD and PLA are characterized by an extended glass transition phase, after which the material begins to flow. On the other hand, the TPU polymer is characterized by a glass transition phase in which both the $\tan\delta$ and E' modulus show a high temperature dependence at the lowest temperature in the samples tested. In PET, as well as the ABS, in both high-viscoelastic deformation phases, there is no temperature dependence of E' and $\tan\delta$ is decreasing to the lowest level showing high temperature dependence. Analyzing the changes in the tangent value of the mechanical loss angle, it was found that the highest stiffness and damping ability is found in ABS in comparison to the remaining samples tested. The smallest stiffness and low vibration damping capacity is characterized by PLA. Moreover, ABS obtained from injection seems to act more stable when we heat it, but ABS samples obtained from 3D printing have a better vibration dumping ability than the injected samples. 3D printing also has some disadvantages such as: long production time comparing to the injection time, printed elements have pores that allow air and water to get through, it is impossible to get sharp angles because printers are limited by nozzle dimensions, when angle is lower than 20 degrees, the printer will cut an area at end of angles due to minimal wall

thickness, etc. Moreover, some of the printing filaments release intense chemical vapours that are harmful.

5 REFERENCES

- [1] Siemiński, P. & Budzik, G. (2015). *Incremental techniques. 3D printing. 3D printers*, Warsaw, WNT.
- [2] Štefanić, I., Raos, P., Samardžić, I., Tintor, B., & Musser, E. (2012). Rapid prototyping of casting cores. *Tehnički vjesnik*, 19(2), 459-464.
- [3] Kaziunas France, A. (2013). *Make: 3D Printing: The Essential Guide to 3D Printers*, Sebastopol, Maker Media.
- [4] Barnatt, Ch. (2014). *3D Printing*, Createspace.
- [5] Wootthikanokkhan, J., Cheachun, T., & Sombatsompop, N. (2013). Crystallization and thermomechanical properties of PLA composites: Effects of additive types and heat treatment. *Journal Applied Polymer Science*, 129, 215-223. <https://doi.org/10.1002/app.38715>
- [6] Vaezi, M. & Chua, C. K. (2011). Effects of layer thickness and binder saturation level parameters on 3D printing process. *Advanced Manufacturing Technology*, 53, 275-284. <https://doi.org/10.1007/s00170-010-2821-1>
- [7] Zemcik, O. & Sedlak, J. (2019). Application of Linear Optimization on Parameters of 3D FDM Print. *Tehnički vjesnik*, 26(4), 1164-1170. <https://doi.org/10.17559/TV-20170506140109>
- [8] Gajdoš, I. & Slota, J. (2013). Influence of printing conditions on structure in FDM prototypes. *Tehnički vjesnik*, 20(2), 231-236.
- [9] Stephens, P. A., Azimi, P., El Orch, Z., & Ramos, T. (2013). Ultrafine Particle Emissions from Desktop 3D Printers, *Atmospheric Environment*, 79, 334-339. <https://doi.org/10.1016/j.atmosenv.2013.06.050>
- [10] Gong, Y. & Guisheng, Y. (2009). Manufacturing and physical properties of all-polyamide composites. *Journal of Materials Science*, 44(17), 4639-4644. <https://doi.org/10.1007/s10853-009-3708-0>
- [11] Galeta, T., Šimunović, G., & Mazurek, M. (2015). Impact of strengthening fluids on roughness of 3D printed models. *Metallurgy*, 24(1), 231-234.
- [12] Sikora, P., Gnatowski, A., & Gołębski, R. (2019) Tests of Mechanical Properties of Semicrystalline and Amorphous Polymeric Materials Produced by 3D Printing. *MATEC Web of Conferences*, 254. <https://doi.org/10.1051/mateconf/201925406003>
- [13] Gnatowski, A. & Chyra, M. (2016). Examinations of Thermomechanical Properties and Structure of Polyethylene with Cardboard Fibre and Colourant Composites after Electrochemical Ageing. *Materials Engineering*, 3, 120-125. <https://doi.org/10.15199/28.2016.3.5>
- [14] Pavlic, T. & Golubic, S. (2016), Analysis of the influence of 3D printing on hardness of parts. *Tehnički vjesnik*, 23(3), 843-847. <https://doi.org/10.17559/TV-20140901165134>
- [15] Tran, J. T. (2015). The Law and 3D Printing. *John Marshall Journal of Computer & Information Law*, 31(4), 505-520.
- [16] Gnatowski, A. (2010). Thermal Analysis of Dynamic Mechanical Properties and Research of the Structure and Crystallinities of Ultraviolet Radiation Ageing Polypropylene with the Dye Addiction. *Composites Theory and Practice*, 4, 328-332.
- [17] Thomas, A. (2008). The 3D Trainwreck: How 3D Printing Will Shake Up Manufacturing. *Analogue*, 128(11), 50-63.
- [18] Juric, A., Štefic, T., & Arbanas, Z. (2011). Experimental analysis of the strength of a polymer produced from recycled material. *Tehnički vjesnik*, 18(4), 627-631.

- [19] Kumar, R., Singh, R., & Farina, I. (2018). On the 3D printing of recycled ABS, PLA and HIPS thermoplastics for structural applications. *PSU Research Review*. 2(2), 115-137
<https://doi.org/10.1108/PRR-07-2018-0018>

Authors' contacts:

Adam GNATOWSKI, Dr. Eng., Assoc. Prof.
Czestochowa University of Technology
Department of Technology and Automation
Al. Armii Krajowej 21, 42-201 Czestochowa, Poland
E-mail: gnatowski@ipp.pcz.pl

Rafał GOŁĘBSKI, Dr. Eng.
Czestochowa University of Technology,
Department of Technology and Automation,
Al. Armii Krajowej 21, 42-201 Czestochowa, Poland
E-mail: rafal@itm.pcz.pl

Piotr SIKORA, MSc
Czestochowa University of Technology,
Department of Technology and Automation,
Al. Armii Krajowej 21, 42-201 Czestochowa, Poland
E-mail: piotrek997@gmail.com

INDUSTRY 4.0 AND BUSINESS PROCESS MANAGEMENT

Jiri TUPA, Frantisek STEINER

Abstract: Production companies are adopting new methods for the improvement of their managing production processes. The cost, quality and time are still key important attributes for process improvement. This is reason why the Business Process Management (BPM) is the core of management in different areas, e.g., quality management, sustainability management. The new technologies based on Internet of Things and Services, SMART solutions, and the concept Industry 4.0 are opening new possibilities of BPM implementation. This theoretical paper deals with a review of new trends in BPM and a presentation of the possibilities of core technologies for Industry 4.0 in phases of BPM applications.

Keywords: Business Process Management; improvement; Industry 4.0; performance measurement; Risk Management

1 INTRODUCTION

Industry 4.0 is a recent keyword in the manufacturing area. The term 'Industry 4.0' has been used mainly by the German government in high-tech strategy. This strategy, with its related current literature, is called Industry 4.0, like '4th Industrial Revolution'.

The increasing number of papers is evidence that this topic is the starting object of research at many research centres and universities. Governments in different countries have adopted concepts that support implementation of the concept Industry 4.0. Concretely the Czech government approved the document 'Initiative Industry 4.0' and allocated support for relevant research projects for examples.

Business Process Management (BPM) is a fundamental theory adopted in large numbers of companies in different industrial sectors. The main ideas are based on the theory of Hammer and Champy [1]. The traditional implementation of this concept is a focus on process analysis, process modelling and optimisation, process automation and measurement of process indicators in relation with performance indicators. The concept BPM has been adopted into international standards, the ISO 9001 for quality management for example.

The concept Industry 4.0 opens new opportunities and research questions. One research question is linked with the future of process management and its implementation for the concept Industry 4.0. This paper tries to find an answer for this research question based on a review of suitable technologies and methods for BPM implementation.

2 LITERATURE REVIEW

BPM is term used in many companies. Over the last decade, BPM has played an important role. BPM helps implement methods and tools for process improving. The BPM principles that combine selected findings from management issue and related management science, information technology [2].

Many growing companies implemented this management approach established on Hammer's Business Process Reengineering Concept [1]. The authors follow up Hammer and Champy's ideas in related research and

practical papers nowadays. Researchers and practitioners from academy and industry use the term BPM in many different ways.

Prof. August-Wilhelm Scheer mentioned in his book that process management has become an important management instrument in many companies because its practitioners realised that optimising a company's capacity to create value in the form of processes has a direct and immediate effect on the bottom line. [3]

The term BPM has been used in many ways, and several alternatives — from process improvement to intelligent BPM (iBPM) — have been proposed. Some use BPM to refer a general theory to the management of business process change, while others use it more narrowly, to refer to the use of automated software techniques to control the runtime execution of enterprises process. Surveys presented on the state of the BPM market in 2018 presents examples of how organisations use the term BPM for example [4].

Different articles in international conference proceedings and books deal with BPM issues [5]. The high number of citations for term business process management (BPM) seems to prove that BPM is an important field of recent research activities. The BPM issue is the point of research focused on technological or methodological solutions of problems related with BPM implementation [6].

Nowadays, the issue focused on BPM is an integrated and continuous process, which deals equally with technological and organisational considerations. It is important to understand that BPM itself represents a process. It is a process consisting of the phases of strategy definition, process description, process implementation and process execution. The whole implementation of BPM includes these important activities:

- Analysis of processes
- Definition of structure between processes
- Choice of management method
- Modelling and optimising of processes
- Determination of performance measurement and diagnostics system.

The main purpose of this process implementation is to improve an organisation's effectiveness in achieving its defined objectives.

The benefits of Business Process Management are to improve the alignment of all aspects of an organisation with internal rules and transparency, to increase customer satisfactory, to promote business efficiency, to maximally exploit employee potential, to improve product quality, and to reduce costs.

Examples of BPM are present in areas where effective progress has been achieved, including, as presented in [5]:

- The systematic validation of business process models before implementing them via information system, to avoid potentially costly mistakes at run on-time.
- The automatic generate of configurable process models from a collection of process model variants, used to guide analysts when selecting the right configuration.
- The complex identification of process behaviours based on scientific insights provided by the workflow patterns initiative.
- The implementation of processes on-the-fly and the evaluation of the impact of their changes, in order to react to (unexpected) exceptions.
- The automatic execution of business process models based on strictly defined semantics, through a variety of information systems supported BPM.
- The automatic finding of process models from raw event data stored in databases of information systems.

2.1 Industry 4.0

The keyword Industry 4.0 is very often cited in recent journal and conference papers. Different authors present this phenomenon like the fourth industrial revolution. Industry 4.0 is a recent trend in the manufacturing world. The term 'Industry 4.0' has been used mainly by the German government in the high-tech-strategy. Historical and current phases are presented in Tab. 1. The table mentions key inventions and technology that are typical for these periods.

Table 1 Core technologies of Industry 4.0 – adopted from [10]

	Phase of Industrial Revolution	Key Invention	Technologies and Capabilities
History	Industry 1.0	Steam machine	Introduction of mechanical production facilities with the help of water and steam power.
	Industry 2.0	Electrical machines and drivers	Introduction of the division of labour and mass production with the help of electricity energy.
	Industry 3.0	Computer (programmable logic controller)	Use of electronic and IT systems that further automate production.
Present	Industry 4.0	Internet	Use of cyber-physical system.

The main idea of Industry 4.0 is the emergence of "smart factories", which will be connected to the production facilities' cyber-physical systems called CPS.

Developing the Internet of Things, the Internet of Services, and the Internet of People will help to make connections between human-machine, human-human or machine-machine, and with it also an enormous amount of data is obtained. For this reason, it will be useful to analyse a large amount of data (Big Data). It be able to predict possible faults and making decision on-line to the changed conditions [7] and [8], for example.

2.2 Policies to Support Industry 4.0 Developments

In its Industrial Policy Papers of 2012, the European Commission defined six basic priorities, three of which thematically cover areas related to the Industry 4.0 concept. Specifically, these include advanced manufacturing, key enabling technologies (e.g. batteries, smart materials and high-performance manufacturing processes), and smart networks and digital infrastructures (where infrastructure and networking are seen as crucial). [10].

Subsequently, in 2013, the Commission appointed a Working Group on Modern Net Production and in 2014 a Strategic Policy Forum on Digital Business to Digital Transform the European Industry and Enterprises. The Forum recommended setting national targets, setting up centers of excellence, strengthening standardization, ensuring funding and encouraging the development of digital skills. European authorities also called for a "observatory" for the digital transition and increasing political awareness about digital issues and policy implications. [10]

In the 2014 declaration "For the European Industrial Renaissance", the European Commission stressed that digital technologies (including cloud computing, new industrial Internet applications, large data, smart factories, robotics and 3D printing) are important to boosting European productivity and innovation in companies through new business models and developing new services and products. [11]

The issue of digital transformation of Europe is one of the most important key actions of the EU. This business and trade transformation is a challenge for Europe and represents a huge growth potential for innovation. Traditional industries in Europe can build on their strengths in advanced digital technologies. This will open up these sectors to gain market share in the emerging markets for products and services of the future.

2.3 Digital Transformation

Digital transformation is defined in the literature as the integration of advanced technologies with physical and digital systems, most innovative business models and new processes, and the development of SMART products and services. [12]

The published Digital Transformation Monitor is the one of the EU action publication. The monitor gives the opportunity to monitor statistical and factual evidence of the pace of digital transformation in specific sectors and in all

EU countries. It tries to identify key trends in digital transformation and to measure progress made at national and sector-specific levels. It also supports policy development and analyses major national policy initiatives. [12]

On the other hand, measurement of digital performance and tracking the development EU member states in digital competitiveness are made by the Digital Economy and Society Index (DESI). The DESI overall index is calculated as the weighted average of the five main DESI dimensions with the weights selected by the user: 1 Connectivity, 2 Human Capital, 3 Use of Internet Services, 4 Integration of Digital Technology, and 5 Digital Public Services. [13]

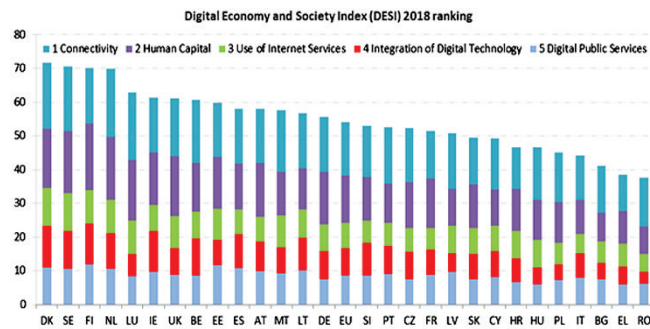


Figure 1 Digital Economy and Society Index (DESI) 2018 ranking [13]

Based on the published assessment in 2018, it is clear that Denmark, Sweden, Finland and the Netherlands are among the most advanced digital economies in the EU, followed by Luxembourg, Ireland, the United Kingdom, Belgium and Estonia (Fig. 1). On the other hand, Romania, Greece and Italy have the lowest scores on DESI. This evaluation can be used very well to compare EU Member States and evaluate their preparation for the implementation of the Industry 4.0 concept. Finally, based on this state of art, we can conclude that the concept Industry 4.0 opens a new dimension of management science.



Figure 2 PDCA model

3 INDUSTRY 4.0 AND BUSINESS PROCESS MANAGEMENT

The number of automated process will be increasing, and due to this fact, traditional management tools and technique should be revised. Business process management is one of

them. Based on this fact, we have tried to answer the research question — What is role of process management for Industry 4.0? — in this paper. The revision of the BPM approach in context of Industry 4.0 has been examined through the Deming PDCA cycle because this cycle describes phases of process BPM implementation. [14]. Fig. 2 presents activities for implementing each phase of PDCA.

3.1 Process Design and Modelling

Process design and modelling are the initial activities of process management implementation. The aim of this activity is to find and describe existing processes and suggest a "to be" state. Areas of focus include process representation, attributes, such as inputs, outputs, resources, and related documentation.

Process modelling is a useful tool to capture, formalise and structure the knowledge of business processes. The models can be deterministic, stochastic, uncertain but structured, or uncertain but imperfectly structured. Modelling tools can be categorised into business process chain-based type such as Petri net, IDEF0, IDEF3, BPMN 2.0, ARIS-eEPC, etc.; formal description type such as WPD, XML, process algebra, predicative logic, etc.; and object-oriented type such as IDEF4 and UML. The overall model is analysed on the individual models. Models would have to contain all process attributes and static and dynamic parameters such as process time, process threats, cost, etc.

New technologies of Industry 4.0 have new role of process design and process modelling. The new role is related with process digitalisation. The term of digitalisation of processes is related with smart factory or digital supply chain network. In the context of process management, digital transformation refers to transforming the business operations, services and models. Digital transformation covers all process in the companies, and the aim is to build a digital model of enterprise with the digitised process attributes (input, output, sources and indicators). The companies should determine rules, methods and standards for process transformation. The result of this step is a digital model of enterprise.

This model contains:

- Relationship between cyber and physical layer
- Description of the main processes (processes that create added value)
- Definition of support processes and autonomous control processes.

3.2 Process Execution

Execution means that instances of a process are performed or enacted, which may include automated aspects. Automation of business processes is based on using the Business Process Execution Language (BPEL). The BPEL language is perfectly suited to a clear description of the processes by which processes can be then mechanically exercised. BPEL enables organisations to automate their processes (so-called service orchestration). The BPEL language allows description of the conduct and behaviour of

business processes as a sequence of activities, including their branches, which are engaged in the process, while these activities are represented by web services (WS) [2].

Business process execution in terms of Industry 4.0 is related with the digital factory and application of Cyber Physical System (CPS). The rules and process model can be used for the definition of elementary relationships in CPS. The cyber level of CPS enables to simulate the process using digital twins and makes an environment for process execution [15].

The CPS implements these elements:

- Data acquisition and data processing
- Machine-to-machine communication
- Human-machine interaction.

3.3 Process Monitoring

The main principles of a process monitoring system are described by many authors, and a lot of them discuss the phrase ‘process performance measurement’. Process performance measurement is the combination of processes, methodologies, metrics and technologies to measure, monitor and manage the performance of the process. To measure and execute process performance, we can use different Key Performance Indicators (KPIs).

These KPIs, also known as process performance indicators, are indicators that can be unambiguously assigned and used for a given process, provided that data is available to calculate these performance indicators [16]. Basic performance parameters are defined, which are time, quality and cost. Specifically, for example, we can measure implementation costs, resource use and waste; time using cycle time, waiting time or no value added time; and quality through customer satisfaction and error rates. Some KPIs can be measured relatively easily, such as cycle time.

The implementation of IoT or Internet of Services, Process and People (Internet of Everthing) and related technologies helps to implement a system for automated and real-time process monitoring. These technologies help to display the state of processes. The automated system of measurement enables data storage in data warehouses, and we are speaking about the analysis of Big Data.

Big data analysis can be defined as the process of examining large and diverse data, identifying hidden patterns, unknown correlations, trends, and other useful information that can help organizations improve their decision-making processes at all levels of management.

Data can be used for:

- Dynamic organisational analysis
- Process optimization
- Right-time monitoring
- Process mining for automated weak point analysis
- Process simulation.

The role of process monitoring is a key process in the CPS. Real-time automated processes can be monitored on-line with the application of a system based on IoT and IoE. The paper [17] presents a model for performance process measurement based on an implementation of IoT

technologies. This model is based on the standards ISA-95 and ISO 22400.

The standards formally describe how a manufacturing process can be formalised and the steps for set-up of the suitable performance indicators. The ISO 22400 standard is a four-part standards; two parts are under development, and two parts are under consideration. The standard ISO 24000-2 defines thirty-four KPIs (key performance indicators) and their equations. Fig. 3, adopted from [17], presents the IoT-based production performance model developed in accordance with ISO 22400 and ISA 95.

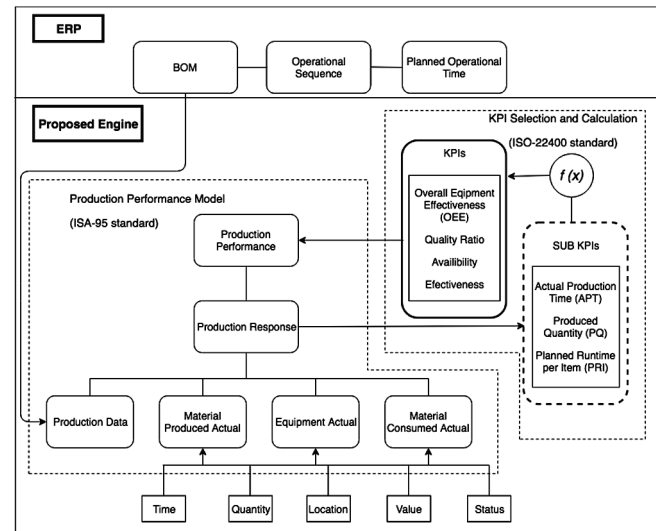


Figure 3 IoT-based production performance model [17]

The model presents data transfer between the production line and the Manufacturing Execution System. Based on the IoT functionality, process data that include information about the time, location, cost, quantity and status can be acquired from the product. The stored data are then adjusted with the production performance model, which includes of material consumed actual, of the equipment actual, material produced actual and of three subparts for example.

IoT data cannot be used directly for all KPIs; then you need to define additional databases for additional data. For example, these are production data, which are ordered according to the sequence of processes, e.g. with the planned process time, the BOM. Production data is obtained from ERP and includes structured information on the production plan. Using the planned and actual data, the above sections classify IoT data based on their description to provide a top-level source for calculating sub-KPIs. The production reaction then uses the sub-KPI to calculate the final KPIs.

3.4 Process Improvement

Process improvement is the last, important phase of business process management. The aim is continual improvement of business processes based on monitoring and data collection. Especially in the context of digital processes, the identification of existing pain points can be realised through the performance of continuous software tests.

Another definition explains process improvement as the proactive task of identifying, analysing and improving upon existing business processes within an organisation for optimisation and meeting new quotas or standards of quality. The companies collect data, not only from manufacturing processes, but from all processes during the product life cycle.

The companies use several methodologies, like Six Sigma, Centric leverages DMAIC (Define, Measure, Analyse, Improve, Control), Lean principles, etc. New possibilities of technologies of Industry 4.0 can effectively support the implementation of the mentioned methodologies. The selection of a continuous improvement methodology depends on the overall strategy. The demonstration of Lean principles and other improvement methodology implementations have been presented in papers [18].

3.5 Discussion

The implementation of Business Process Management for the concept Industry 4.0 is a new topic, which is not exactly mentioned in the literature. The SWOT analysis (Tab. 2) was used for a summary of the strength and weakness and the opportunities and threats for a company based on previous subchapters (3.1–3.4) and related literature.

Table 2 SWOT analysis

Strengths: <ul style="list-style-type: none"> Theoretical background Knowledge from Business Process Management Existing standards for process digitalisation Using Big Data for process monitoring, controlling and performance management 	Weaknesses <ul style="list-style-type: none"> Cost for implementation of CPS or IoT Training of staff and personal Time to implementation No existing related suitable methods and tools for process optimisation
Opportunities: <ul style="list-style-type: none"> Develop new lead markets for products and services Open a new challenge for markets Development of effective supply chains 	Threats <ul style="list-style-type: none"> Cybersecurity and IT criminality Lack of IPR protection Vulnerability to volatility of economy

As mentioned, the important part of business process management is to define the suitable architecture and rules which help to implement the concept Industry 4.0.

4 ARCHITECTURE AND RULES FOR BUSINESS PROCESS MANAGEMENT IMPLEMENTATION

Architecture and rules provide a common structure and language for the uniform description and specification of system architectures. The concept Industry 4.0 uses a reference architecture model. Reference Architectural Model Industry 4.0, abbreviated RAMI 4.0 (Fig. 4), consists of a three-dimensional coordinate system that describes all crucial aspects of Industry 4.0. Indicated on the right horizontal axis are hierarchy levels from IEC 62264, the international standards series for enterprise IT and control systems. These hierarchy levels represent the different

functionalities within factories or facilities. The left horizontal axis represents the life cycle of facilities and products, based on IEC 62890 for life-cycle management. [19]

The six layers on the vertical axis serve to describe the decomposition of a machine into its properties structured layer by layer, i.e. the virtual mapping of a machine. The RAMI architecture links to other important standards for concept Industry 4.0 implementation.

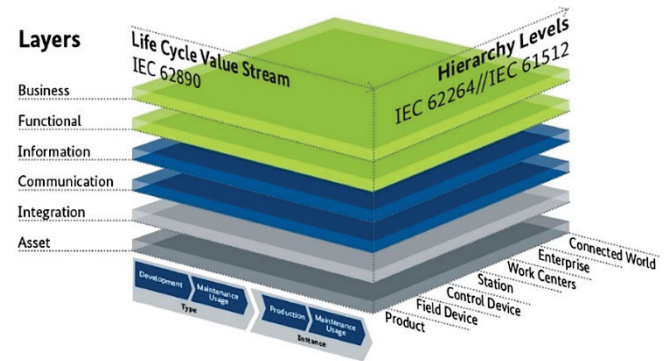


Figure 4 Reference Architectural Model Industry 4.0 (RAMI)

The Tab. 3 presents areas of implementation and titles of selected standards [20]. The architecture RAMI and related standards help to analyse and describe basic processes at the smart factory and all processes related with digital supply chain. All the business processes can be adopted in this model, and this model can set up business rules and architecture for a smart factory.

Table 3 Standards for Industry 4.0 [20]

Area of implementation	Title of standard
Security	Information Security Management Systems, Network and System Security
Safety	Functional Safety, Safety of machinery
Digital factory	Reference model for the digital factory
Information layers	Classification and product description
Configuration	Function blocks for process control/electronic device description language Field device tool (FDT) interface specification
Hierarchy level	Batch control, Enterprise-control system integration
Communication layer	Industrial communication network, Machine-to-machine communication
Life cycle	Life-cycle status
Condition	Condition monitoring
Energy	Energy efficiency
Engineering	Industrial automation systems and integration, Automation Markup Language
Semantics	W3C Semantic Web stack
Management	Quality management, Risk Management, Key Performance Indicators for Manufacturing Operations

5 RISK MANAGEMENT

Risk management is an important part in the business process management area nowadays. Enterprises try to implement risk management procedures as part of the

management process. They try to implement management system with process, quality and risk management for performance measurement. This way is also recommended in ISO standards for quality management system described in the new revision of ISO 9000 standards in year 2015.

The task of the risk management is a systematic approach to assessing and acting on risks in order to ensure that organisational objectives are achieved. The concept Industry 4.0 opens a new challenge for risk management nowadays.

The mentioned concept Industry 4.0 can generate new kinds of risks in related area because of the increase of threats and vulnerability. The networking of cyber-space, advanced manufacturing technologies and elements, and implementation of services outsourcing is the main factor increasing the vulnerability. An identification of new categories of risks is presented in Tab. 4.

Table 4 Identification of new risks adopted from [21]

Categories of manufacturing risks	Risks
Maintenance	Problem with availability and integrity of data for maintenance.
Used manufacturing methods and tools	Errors in data processing.
Manufacturing process management	Information risk associated with data losses, break of integrity and available information.
Human sources	Low number of qualified workers
Machine environments	Attacks from Internet network, problems related to electromagnetic compatibility and electromagnetic emissions affect manufacturing machines.
Machines and manufacturing technologies	Sensitivity and vulnerability of data—problem related to cyber-attacks.

This identification has recommend an important steps for risk management implementation.

The results of our finding present that the most of common risk factors in the manufacturing area can be associated to information security. A discussion about a new term — Information, Communication and Manufacturing Technologies (ICMT) — which reflects the requirements of Industry 4.0 can be started. The reason is that manufacturing infrastructure autonomous intelligent manufacturing system, machines and robots communicate with each other, which should be included under the umbrella of ICT. The important question is how to protect the ICMT against cyber-attacks, loss of data integrity or problems with availability of information. An implementation of information security management systems can answer this question.

6 CONCLUSION

The aim of this theoretical contribution is to summarise the key elements of Business Process Management for the adoption of the concept Industry 4.0. The concept Industry 4.0 opens new challenges for process management implementation nowadays. The technologies based on CPS, IoT, Digital Factory etc. help to improve processes based on

effective process measurement, simulation and application of predictive models.

The implementation of process management has been demonstrated by using the PDCA cycle. The core of this concept implementation is based on different rules and standards integrated in the RAMI model.

The implementation of BPM is associated with the risk management area. In this point of view, the concept Industry 4.0 can generate new risks in the manufacturing area. These risks are associated with cyber-attacks, for example, the loss of data integrity etc. The identification of these new risks was described in Chapter 4, and the chapter answered the first research question: What new kinds of risks are associated with manufacturing concept Industry 4.0? In the face of new threats, companies should try to implement an information security management system into the manufacturing area

We conclude that Business Process Management is the perspective discipline nowadays, and with the concept Industry 4.0, new challenges for its successful implementation are opened.

Acknowledgements

This research has been supported by the Ministry of Education, Youth and Sports of the Czech Republic under the RICE – New Technologies and Concepts for Smart Industrial Systems, project No. LO1607 and by the Technology Agency of the Czech Republic under the project Software platform to accelerate the implementation of management systems and process automation — project No. TH02010577.

7 REFERENCES

- [1] Hammer, M. & Champy, J. (1993). *Business process re-engineering*. London: Nicholas Brealey.
- [2] Weske, M. (2012). *Business process management: Concepts, languages, architectures*. Berlin: Springer.
<https://doi.org/10.1007/978-3-642-28616-2>
- [3] Scheer, A., Abolhassan, F., Jost, W., & Kirchmer, M. (Eds.). (2002). *Business process excellence*. Berlin: Springer.
<https://doi.org/10.1007/978-3-540-24705-0>
- [4] Harmon, P. (2018). *The state of business process management – 2018*. BPTrends.
- [5] van der Aalst, W. M. P., La Rosa, M., & Santoro, F. M. (2016). Business process management. *Business & Information Systems Engineering*, 58(1), 1-6.
<https://doi.org/10.1007/s12599-015-0409-x>
- [6] Jeston, J. & Nelis, J. (2008). *Business process management*. Oxford: Butterworth-Heinemann.
<https://doi.org/10.4324/9780080492520>
- [7] Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014) Industry 4.0. *Business & Information Systems Engineering*, 6(4), 239-242.
<https://doi.org/10.1007/s12599-014-0334-4>
- [8] Richert, A., Shehadeh, M., Plumanns, L., Gros, K., Schuster, K., & Jeschke, S. (2016). Educating engineers for industry 4.0: Virtual worlds and human-robot-teams: Empirical studies towards a new educational age. In *IEEE Global Engineering Education Conference (EDUCON)*, 142-149.
<https://doi.org/10.1109/EDUCON.2016.7474545>
- [9] Lopes de Sousa Jabbour, A. B., Jabbour, C. J. C., Godinho

- Filho, M., & Roubaud, D. (2018). Industry 4.0 and the circular economy: A proposed research agenda and original roadmap for sustainable operations. *Annals of Operations Research*, 270, 273-286. <https://doi.org/10.1007/s10479-018-2772-8>
- [10] Davies, R. (2015) *Industry 4.0. Digitalisation for productivity and growth*. European Parliamentary Research Service.
- [11] European Commission (2014). Towards an Industrial Renaissance. Retrieved from https://ec.europa.eu/growth/industry/policy/renaissance_en
- [12] European Commission (2017). Digital transformation. Retrieved from https://ec.europa.eu/growth/industry/policy/digital-transformation_en
- [13] European Commission (2019). The Digital Economy and Society Index (DESI). Retrieved from <https://ec.europa.eu/digital-single-market/en/desi>
- [14] Aziz, N. A. A., Manab, N. A., & Othman, S. N. (2016). Critical success factors of sustainability risk management (SRM) practices in Malaysian environmentally sensitive industries. *Procedia - Social and Behavioral Sciences*, 219, 4-11. <https://doi.org/10.1016/j.sbspro.2016.04.025>
- [15] Prause, G. (2015). Sustainable business models and structures for Industry 4.0. *Journal of Security & Sustainability Issues*, 5(2), 159-169. [https://doi.org/10.9770/jssi.2015.5.2\(3\)](https://doi.org/10.9770/jssi.2015.5.2(3))
- [16] Gündüz, M. (2015). Value stream performance measurement in lean manufacturing business. *International Business and Management*, 10(3), 40-47.
- [17] Hwang, G., Lee, J., Park, J., & Chang, T.-W. (2017). Developing performance measurement system for Internet of Things and smart factory environment. *International Journal of Production Research*, 55(9), 2590-2602. <https://doi.org/10.1080/00207543.2016.1245883>
- [18] Wagner, T., Herrmann, C., & Thiede, S. (2017). Industry 4.0 impacts on lean production systems. *Procedia CIRP*, 63, 125-131. <https://doi.org/10.1016/j.procir.2017.02.041>
- [19] Schweichhart, K. (2016). *Reference Architectural Model Industrie 4.0*. Berlin: Federal Ministry for Economic Affairs and Energy.
- [20] Enterprise Information Systems (2017). Standards for Industry 4.0. [Online]. Retrieved from <http://i40.semantic-interoperability.org/>
- [21] Tupa, J., Simota, J., & Steiner, F. (2017). Aspects of Risk Management Implementation for Industry 4.0 in Standards for Industry 4.0. *Procedia Manufacturing*, 11(2017), 1223-1230. <https://doi.org/10.1016/j.promfg.2017.07.248>

Authors' contacts:

Jiri TUPA, Corresponding author
University of West Bohemia in Pilsen
Faculty of Electrical Engineering
Univerzitni 8, 306 14 Pilsen, Czech Republic
tupa@ket.zcu.cz

Frantisek STEINER
University of West Bohemia in Pilsen
Faculty of Electrical Engineering
Univerzitni 8, 306 14 Pilsen, Czech Republic

LEADERSHIP AND CONFLICT MANAGEMENT IN PRODUCTION ORGANISATIONS

Jerko GLAVAŠ, Veljko KONDIĆ, Iva BALIĆ

Abstract: Conflicts between different interests are inevitable in every business organization. Therefore, it is not surprising that the number and intensity of conflicts is sometimes very high. Conflict resolution is a skill without which managers can hardly succeed in their business. The subject of research in this paper is conflict management from a technical and economic point of view. The paper defines and analyzes the notion of leadership, that is, elements and dimensions of leadership, the notion of leader, and theoretical approaches to leadership. In addition to the above, the concept of conflict, more specifically its type, levels of conflict, consequences and causes, and techniques of conflict resolution have been defined and analyzed. An online survey (based on the LinkedIn Business Social Network) was conducted to determine how individuals in a production organization respond and resolve conflicts as well as the results of the same. Based on the theoretical scientific approach, the conducted research and the analysis of the results, conclusions and recommendations are given, as well as key questions for future research.

Keywords: conflict; leadership; management; production organization

1 INTRODUCTORY CONSIDERATIONS

Leadership is considered to be one of the most important factors for a company's success. The notion of leadership is very complex, which is why there are many variations in defining it. Within this thesis, leadership will be seen as the ability to influence the group in terms of managing conflicts to achieve organizational goals.

In conflict situations, the role of the leader is to direct and motivate employees, all for the purpose of completing their work tasks, in order to contribute to the achievement of the organization's goals through their actions and reflections.

Table 1 Overview of leadership definitions [1]

"Leadership is the ability to influence, inspire and direct individuals or groups toward achieving goals." (Bass, M. B., 1985: 9)
"Leadership is the art or process of influencing people, so that they readily and enthusiastically strive for group goals." (Koontz, H., Weihrich, H., 1994: 437)
"Leadership is the ability to influence, inspire, and direct the actions of individuals or groups toward the desired goals." (Hellriegel, D., Slocum, J. W., 1989: 465)
"Leadership is the process of influencing the activities of individuals or groups towards the achievement of goals." (Gatewood, D. R., Taylor, R. R., Farrell, O. C., 1995: 7)
"Leadership is the ability to influence people toward their goals. Leadership is mutual (reciprocal), it takes place among people. Leadership is a "human" activity, different from administrative paperwork or problem solving activities. Leadership is dynamic and involves the use of power to influence people to move toward goals." (Daft, L. R. 1997: 590)
"Leadership is a process and a characteristic. As a process - focusing on what the leader is currently doing - leadership is the use of non-coercive influence to shape group or organizational goals, motivate behavior toward those goals, and help define a group or organizational culture. As an attribute, leadership is a set of characteristics of individuals who are perceived as leaders." (Griffin, W. R., 2002: 520)

While we define management as the process of directing others toward the accomplishment of a particular task, the emphasis is more on the ability to influence others who perform certain tasks. Leadership is actually the only real function of management because it cannot be delegated to others, while for all other functions special departments can be organized, such as for planning, organizing, human resources management, etc. Leadership is considered one of

the most important factors for the success of a company. The notion of leadership is very complex, which is why there are many variations in defining it. Different views on the conceptual definition of leadership result from the fact that different authors define leadership most often according to their personal views and the viewpoint of the phenomenon that interests them most, resulting in numerous definitions of leadership. The following table shows some of the definitions of leadership by different authors.

Leadership represents the ability to influence a group in terms of achieving goals, and the source of that influence can be formal, such as the influence that ensures a managerial position in an organization. Managerial positions most often come with some degree of formally established authority, which is why individuals may consider that a leadership role derives from the position a manager has in an organization. Not every leader is a manager, nor is every manager a leader. Even though an organization provides certain rights to its managers, this does not mean that managers will be able to lead effectively. Equally important or more important than formal influence is the non-sanctioned leadership, that is, the capacity to influence outside of the formal structure of the organization. In simplified terms, leaders can emerge from the ranks of the group just as they can be formally appointed.

1.1 Leadership as Management Tool

Leadership is generally understood as the most demanding managerial function. One definition is that leadership is actually the art of influencing people so that they willingly and enthusiastically strive to achieve the goals of the organization. Srića [2] defines leadership as "the ability to influence the behavior of associates and their value system to achieve the goals of the organization with enthusiasm". Karlic and Hadelan [3] state that "current knowledge of the leadership process suggests that it is a dynamic process that requires creativity and organization in order to coordinate ongoing business activities while starting new and extinguishing old ones."

Although leadership definitions differ, most contain essential components that are common to all.

Leadership dimensions comprise specific leadership features that make it possible to find, explain, and evaluate differences that exist in leadership content, and are the result of researchers' understanding of the features that make leadership effective. Leadership is a process in which an individual influences a group in order to achieve a common goal. Following the above definition, we see that being a leader means an important interaction with their followers or team members. Transactional leadership is the one in which a group of people is focused on a common goal, although it is not expected of individuals to be proactive, but to perform the tasks that are intended for them. [8]

1.2 Leadership Dimensions and Impact on Conflict

Leadership has been one of the most important variables in interpreting organizational results and the employees' work behaviour. In turn, the classic theories focused either on the characteristics of leaders, on their behaviour, on the environmental factors, or on their interactions with the followers in order to construct interpretations for conduct, efficiency and effectiveness of workers and organizations (Horner, 1997; Van Seters and Field, 1990). Key criteria for this interpretation were the two classic approaches to leadership outlined by the scholars of Ohio State University and the University of Michigan, namely the "consideration, or employee orientation, or peopleoriented leadership" and the "initiation of structure, or production orientation, or taskoriented leadership" (Armandi et al., 2003; Avolio, 2007; Silva, 2015). Bass (1985, as cited in Bass, 1990), based on the emerging concepts of charismatic leadership, initiated by House (1977, as cited in House, 1996) and transformational leadership, devised by Burns (1978, as cited in Yammarino and Bass, 1990) established a new leadership model known as Multifactor Leadership Theory which included the Transformational, Transactional and Laissez-faire Leadership types (Bass, 1990; Bass and Avolio, 1990). Subsequently, Bass and Avolio (1992; 1995) initiated a full research questionnaire named the Multifactor Leadership Questionnaire/MLQ in order to measure and evaluate these new types of leadership and their particular characteristics and effects. [10]

For a long time, it has been a common belief that organizing and operating a business is a rational process in which emotions have no place. Today it is considered that the expression of positive emotions has a very positive effect, which is most pronounced when connected emotionally and rationally, while the neglect of emotions can lead to tragic results. Emotional intelligence is considered the skill of recognizing one's own feelings as well as that of other people, motivating oneself, and managing their emotions and relationships well. Many authors consider emotional intelligence to be more important for high-performance jobs and for the highest levels of leadership than for cognitive and technical skills. Effective leadership requires excellently developed emotional intelligence, and good leaders use their own emotional intelligence rather than positional power.

Emotional intelligence, along with cognitive skills, is crucial in identifying and promoting shareholder value, which supports the execution of vision, mission and strategies.

At the beginning of the twentieth century, when the research of leadership and characteristics of a successful leader began, personal traits were the starting point, as indicated by Yammarino et al.. Cowley very early indicated that it was believed that leaders were born, not made and that it was a matter of genetic predisposition whether someone would become a successful leader or not. Further research raised the question of leadership's behaviour and, therefore, in the second half of the twentieth century, many theories of leadership styles that describe leadership ranging from authoritarian (i.e. task-oriented leadership style) to democratic (i.e. peopleoriented leadership style) appeared.

In addition, the attitude that leaders are born, not made has changed into the attitude that leadership is a skill that can be learned. The continuation of scientific research on leadership showed that the personality of the leader and his behaviour are not the only factors, which have an impact on the successful leadership style, but that it is also necessary to take situational factors into account. In other words, the leader should first define factors that are crucial in a particular situation in order to apply the appropriate leadership style in accordance with requirements of the situation. [14]

Garaca [4] states that the job of leadership has first and foremost an interpersonal dimension, but also a technical one, which means quality, efficiency and productivity.

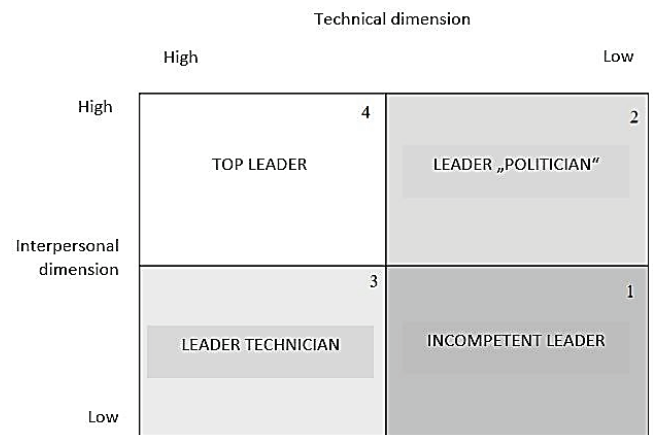


Figure 1 Relationship between the interpersonal and technical dimensions of leadership [4]

The incompetent leader is weak in both technical and interpersonal skills. A politician is a type of leader who lacks technical knowledge, but is good in the interpersonal and social, which is of greater importance when it comes to leadership. The technician is strong in technical knowledge but weak in interpersonal relationships. A top leader is a leader who has all the necessary leadership qualities.

1.3 Conflict as a Business Chance

The issue of conflicts and conflict behavior of the person becomes more and more relevant in different aspects. First of

all, it is connected with the continuing and even increasing conflicts between countries, peoples, ethnic groups, which are clearly observed recently in some regions of the world. In addition, it is associated with a competitive relationship between organizations. Conflicts are an inevitable way of resolving contradictions in any organization. For today's organizations is characterized by a wide variety of conflicts caused by different reasons, they become involved as separate individuals, and groups, driven by different interests and needs that conflict forms the vast space. And naturally this actualizes the issue of studying conflicts and their causes, resolution and management. [11]

A modern trend of study of conflicts is the search for interdisciplinary approaches to understanding the causes and manifestations of conflict behavior in an organization, such as interdisciplinary business psychology approach. This approach combine the psychological and managerial basics. Psychological aspects, are concerned with the nature of conflict behavior and factors of conflict resolution. Conflict is considered as a lack of agreement between two or more parties, individuals or groups, the clash of their opposing interests (Antsupov, Shipilov, 2000 Grishina, 2008). Managerial approach is focused on conflict resolution and conflict management in an organization: on politics as cases give rise to conflict of employee and group of employees (George, Jones, 2007), capabilities of successful organization in the conflict management mechanism for conflict resolution and improving performance (Hart, 2000) etc. That is why the business psychology pay attention on psychological mechanisms and factors of conflict and its importance for the conflict management. [11]

For people in the organization, but also for those who are just watching, either directly or indirectly, conflicts in the organization can be very frustrating. For this other group, it is not very comfortable to work in an environment where two people are constantly fighting and are hostile to one another. Conflicts, by themselves, do not necessarily have to be bad, but have both positive and negative sides. A certain level of conflict within an organization can have positive consequences not only for individuals and groups, but for the organization as a whole. They give people the opportunity to identify problems and opportunities that go unnoticed, and as a result increase creativity and productivity. Successful managers know how to foster constructive conflict in situations where the existing status quo impedes change and development.

"The Positive Consequences of Conflict:

- Improves the quality of decisions made
- Encourages creativity and innovation
- Increases interest and curiosity in the group
- Becomes a mean of reducing tension
- Allows better adaptation to changes.

Negative Consequences of Conflict:

- Dissatisfaction
- Weakening of interconnections
- Decrease in work performance
- Poor communication among members

- Fighting among members becomes more important than doing the job." [5]

When faced with conflict situations, people use one of the following strategies to resolve them:

- "Avoidance is the physical or mental withdrawal from conflict. In this situation, both sides pretend that the conflict does not exist and hope that it will somehow disappear on its own." [5]
- Indulgence is characterized by adjusting to the interests of the other person, that is, one party to the conflict cares more about the other person than about himself, and most often to his detriment. With this technique, people try to reduce the differences between them and emphasize the similarities and what they have in common, which can work for a certain time, but this causes the conflict to be resolved and after a while the conflict may break even more and result in even more frustration and resentment. An individual's behavior can range from immediate behavior to a willingness to relent after a brief reflection on the situation. As stated by Rijavec and Miljković [5], "a person's behavior can act positively, and after a while it may begin to show a so-called passive-aggressive behavior, ie stubbornness, forgetting and the like."
- Imposing is the use of power to accomplish your goals. It is based on aggression and domination, and one person in a conflict thinks only of their desires and completely neglects the other. The imposition can also take the form of command, in such a way that a person with formal power orders another person to do what he has to do. "But in a positive case, this style may include a willingness to understand and respect another person's position. A person wants to explain, prove, and impose his or her attitude, but it also allows the other party to do it." [5] In a competitive environment, people have to fight for respect and influence, which is why they often pretend to be more confident in themselves than they really are. The consequence is that they do not actually dare to ask for the information they need and learn more slowly and do less well.
- Compromise is the search for the most optimal solution or the willingness to give up something in exchange for something else. Compromise means to take care of himself and others without the outcome of gain or loss. Although the conflict may seem resolved, it may re-emerge after a while.

According to Rijavec and Miljković [5], "cooperation is an open and direct confrontation with the conflict and the search for a mutually satisfactory solution." One conflict person is trying to get both people to the maximum and results in a win-win position. Collaboration requires the most precious things in the organization, which are time and energy. Insignificant problems do not require perfect solutions but can be solved in any way since the consequences are irrelevant. Likewise, not every personal conflict is necessarily resolved. Excessive use of a cooperative strategy may indicate that one person in conflict

does not want to take responsibility or that he or she wants to delay something that should actually be done.

2 CONFLICT MANAGEMENT IN ORGANIZATIONS: EMPIRICAL RESEARCH

The aim of the research was to investigate how people respond to and manage workplace conflicts. For the purpose of the research, the authors conducted a survey containing 15 questions with offered answers, of which only one had to be selected. The answers to the questions from line 3 to line 12 have offered answers from 1 to 5. Number 1 represents a situation that never happens, number 2 rarely, number 3 sometimes, number 4 often and number 5 always.

The basic hypothesis: H1: Conflicts in the organization are under the direct supervision of employees and managers. The survey was approached by 82 respondents of different age groups, of whom 56 were female and 26 were male. The survey lasted from August 26 to September 9, 2019 in Osijek. The research is done electronically (via the LinkedIn Business Social Network). The survey included a request to complete the survey and a brief explanation of the survey and its purpose. After the data were collected, these were analyzed by the statistical methods shown below.

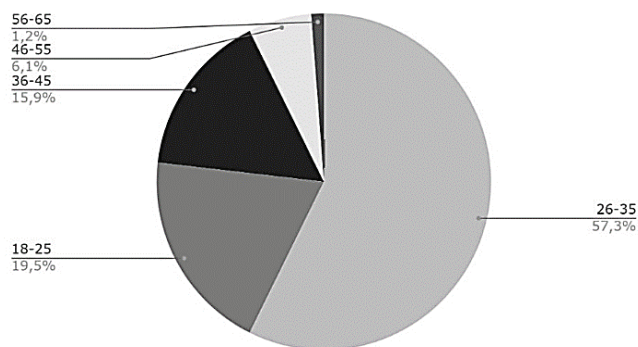


Figure 2 Age of respondents

The graph shows that the age structure of the respondents is different. The majority of respondents were aged 26 to 35 years or 57.3%, followed by respondents between 18 and 25 years or 19.5%. Respondents aged 36-45 accounted for 15.9%, followed by respondents aged 46-55 years or 6.1% and aged 56-65 years or 1.2%.

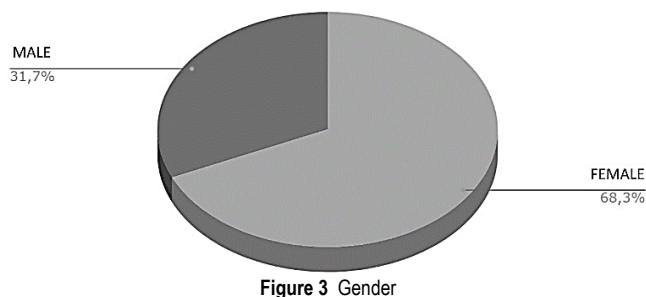


Figure 3 Gender

As mentioned earlier, 55 women or 68.3% of the total surveyed and 26 men or 31.7% of the respondents accessed the survey.

The third question asked was: "It is more important for me in teamwork to be right than to work well with others." The highest number of respondents or 39.02% said that it was rarely important for them to be right, and 35.36% said that it was not important for them to be right at all, but more important for them was to maintain good team relations.

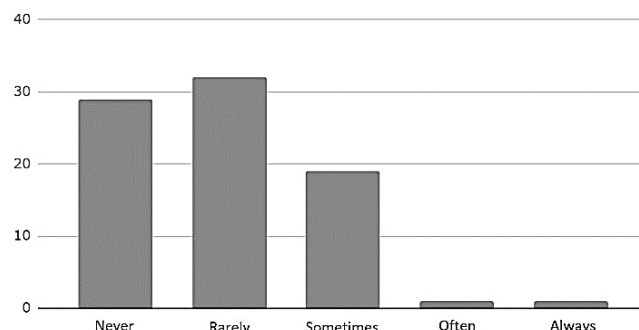


Figure 4 In teamwork, it is more important for me to be right than to work well with others

Furthermore, 23.18% of the respondents answered that it is sometimes more important for them to be right, and 1.22% of respondents must be often or always right.

Fig. 4 refers to the results of the respondents to question 4, which says "I defend my position when someone disagrees with me." According to the data in the graph, 52.44% of the respondents answered the question affirmatively, that is, they always defend their position, and 31.01% do it often.

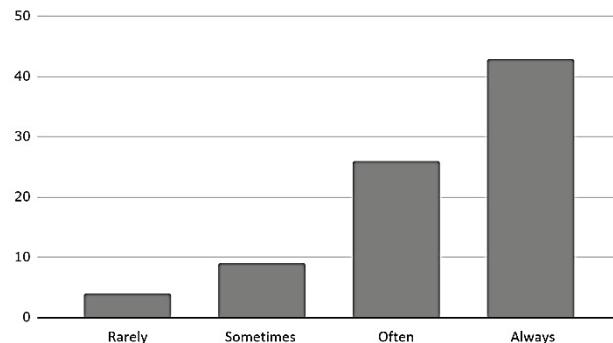


Figure 5 I defend my position when someone disagrees with me

10.97% of respondents answered that they defend their attitude sometimes and 4.87% defend their personal views rarely. None of the respondents answered that they never defend their views.

Fig. 5 refers to how often respondents listen to the advice of others, although they sometimes disagree with them. The graph shows that 41.46% of respondents always listen to the advice of others, and 30.49% of respondents often.

A total of 6.1% of respondents said they rarely or never listen to the advice of others.

The next question asked was "I do not like to be in conflict with others, so I often give in". 1.29% of the surveyed never yield, and 2.44% always yield. Fig. 6 shows the data listed.

Furthermore, 37.80% or the majority of respondents answered that they sometimes give in, and 31.7% often do. 26.83% of respondents rarely give in.

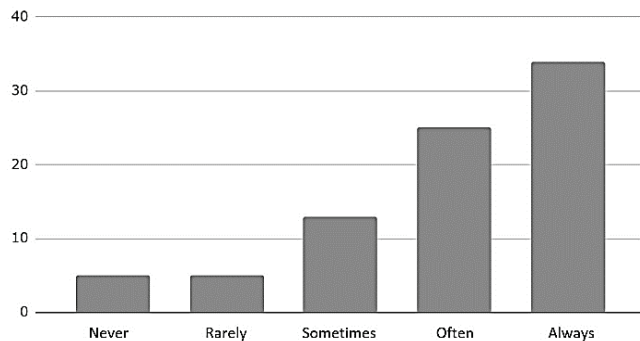


Figure 6 I listen to the advice of others even though I sometimes disagree with them

Fig. 7 gives information regarding whether respondents retain their opinions for themselves not to get in conflict with others. The majority of respondents answered that they rarely keep their opinions to themselves, 32.92% to be exact, and 30.49% always express them.

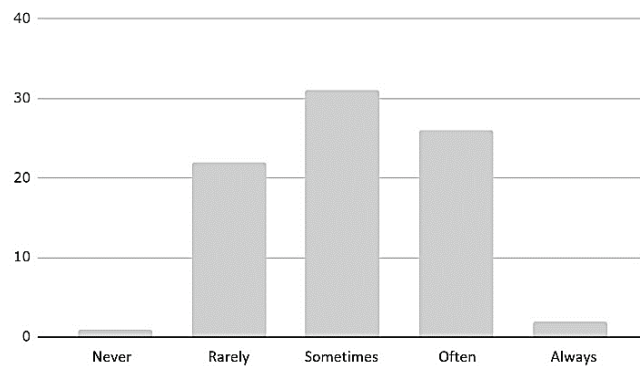


Figure 7 I don't like to be in conflict with others so I often give in

12.19% of respondents sometimes keep their opinions for themselves and 19.51% of the respondents often. 4.87% of respondents always keep their opinions for themselves.

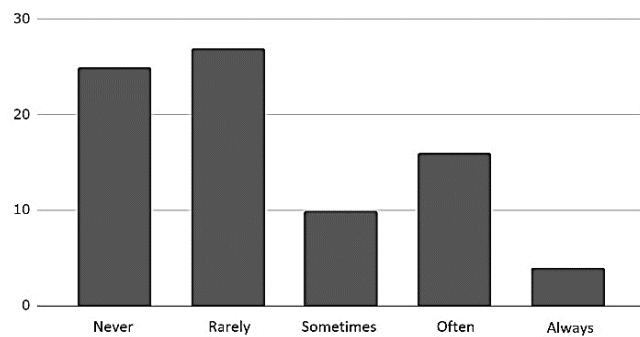


Figure 8 I keep my opinion to myself for fear of conflict with others

Fig. 8 concludes that the largest number of respondents or 48.7% try to get the best out of the conflict for both parties. It is often done by 41.46% respondents. Just 9.76% of respondents sometimes come out of conflict in a win-win situation.

None of the respondents answered that they rarely or never try to bring out the best for both parties. Fig. 9 shows how often people are willing to give in just not to jeopardize the relationship.

None of the respondents answered that they rarely or never try to bring out the best for both parties. Fig. 9 shows how often people are willing to give in just not to jeopardize the relationship.

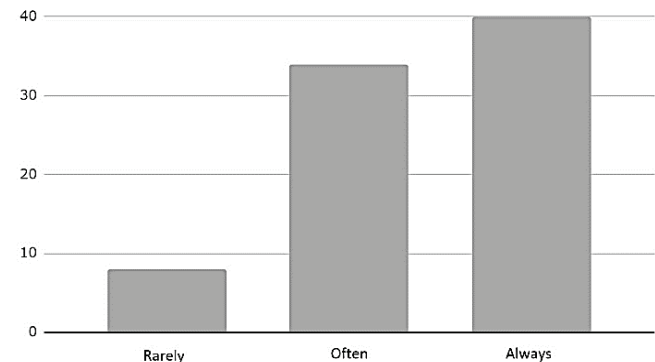


Figure 9 I try to get the best out of the conflict for both sides

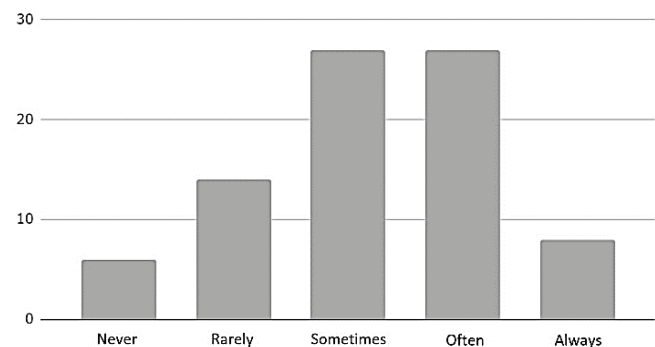


Figure 10 I am ready to give in to others so as not to jeopardize our relationship

The majority of respondents or 32.93%, sometimes or often give in to others. It will rarely be done by 17.07% of respondents, never by 7.31%. A smaller number of respondents or 9.76% will always give in for good workplace relations.

When in conflict, respondents almost always talk to another person to find a common solution to the existing problem.

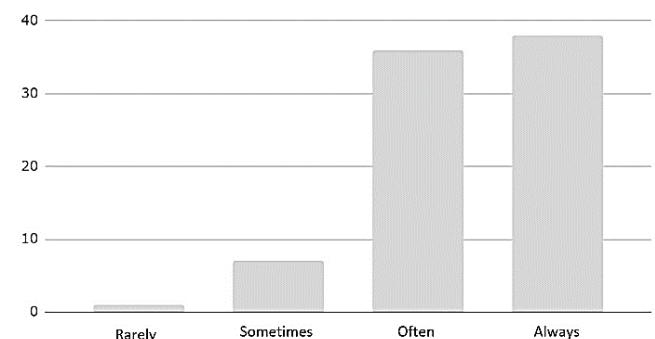


Figure 11 I speak openly with the person I am in conflict to find a solution

According to the survey results, 46.38% of respondents always talk openly with another person, and 43.90% often.

Sometimes the problem is discussed by 8.54%, rarely by 1.22% of the respondents. None of those interviewed answered that they never talk to another person they have a conflict with.

Fig. 11 shows how often the problem with another person is stated to the superior.

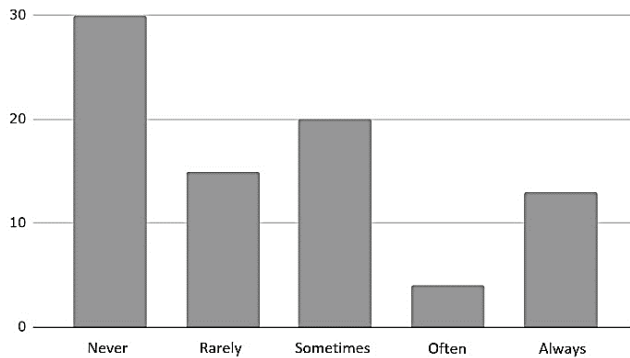


Figure 12 When I have a problem with someone I tell my boss first

A large number of respondents or 36.58% never presents the problem to the boss, but tries to solve it independently. 18.29% of respondents will rarely, and 24.39% of respondents will sometimes address their superiors. 4.88% of the respondents often address their boss and 15.85% always do.

The last question asked in the survey is whether people intentionally provoke others just to cause conflict.

According to the data seen in Fig. 12, it can be concluded that the majority of those surveyed, namely 87.8%, never do so. It will rarely be done by 1.22% of respondents. Sometimes a conflict is provoked by 9.75% and often by 1.22% of respondents.

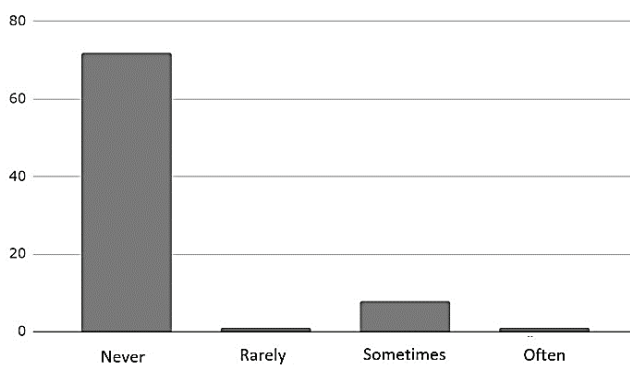


Figure 13 I provoke others just to cause conflict

Based on all of the above, an equation can be defined through the basic components of organizational behavior:

$$\text{ORB ing} = \text{ORB ind} + \text{ORB inv} + \text{ORB before} + \text{ORB conf}$$

ORB - Organizational Behavior

ORB ing - Organizational Behavior (as management engineering)

ORB ind - Organizational Behavior (identifying the behavior of members of the organization)

ORB inv - Organizational Behavior (exploring basic models of organizational behaviour)

ORB pre - Organizational Behavior (predicting the behavior of organization members)

ORB conf - Organizational Behavior (estimation of the level of conflict in the organization).

Estimating and listing these variables can lead to the value of an unknown variable, if necessary, and to the overall state of organizational behavior, based on which conclusions and decisions can be made.

3 CONCLUSION AND RECOMMENDATIONS

Leadership is the only true function of management because it cannot be transferred to others. It is considered to be one of the most important factors for the success of all companies, regardless of their primary activity. [6] For a long time, organizing and operating an enterprise was considered to be a rational process in which emotions have no place, while today it is believed that expressing positive emotions has a very positive effect, which comes to expression when connected emotionally and rationally. The importance of social dimension, ie to understand the human behavior and needs at the workplace for increasing organizational performance was emphasized by behavioral theories according to which participatory management presents the best way to use human resources. Effective leadership requires well-developed emotional intelligence, and every good leader uses his or her own emotional intelligence power rather than positional power.

A conflict is a situation in which two parties, individuals or groups, want to achieve a goal that they believe can be achieved by only one party. [7] Conflicts between different interests are inevitable in every business organization, so it should not be a surprise when the number and intensity of conflicts sometimes are very high. Conflict resolution is a skill without which managers could hardly succeed in their business. A manager must be competent to recognize a situation that can lead to conflict and to deal with that situation in a way that makes employees happy without harming the organization.

In order to answer the question of how people respond to conflict situations in the workplace and how they handle it, a survey was conducted on a sample of 82 people. The results of the survey are in line with the general perception but with some discrepancies, which is evident from the analysis of the results. The hypothesis H1 set in the paper is proven.

It is more important for the outcasts to work with the team than the fact that they are right. Most respondents will defend their position if someone disagrees with them, and most will obey the advice of others even though they disagree. The majority of respondents do not like to be in conflict with others, so for this reason they sometimes relent in conflict situations. Respondents rarely keep their position to themselves for fear of clashing with others, and winning the debate is rarely more important to them than the feelings of others. Most are trying to get the best out of conflict for both sides. For this reason, respondents often adapt to the

demands of others, but they also firmly stand behind their opinions in negotiations. Respondents are sometimes, or often, willing to stumble, so as not to compromise their relationship with others, and always talk to the person they are in conflict with in order to find a common solution. For the most part, respondents do not inform the superior about the conflict first and never provoke others to cause conflict.

From the stated fact that employees do not inform the management about the existing conflict, a key question arises: How to manage conflict in production organizations if we do not know that it exists? It is a key hypothetical question for further research and is the subject of study of numerous systems and activities because it is crucial in building an organizational culture climate that is a prerequisite for success.

In this paper, leadership was seen as the ability to influence the group in terms of managing conflicts in order to achieve organizational goals. For a long time, workplace conflicts were considered a bad thing that should be avoided. However, a conflict-free group is immune to change and innovation. For this reason, a certain amount of conflict within a group is desirable in order to remain creative and self-critical. The role of the leader in conflict situations is crucial, as the leader needs to recognize conflict situations and deal with them through dialogue and compromise, always bearing in mind the primary goal - the well-being of the organization and the achievement of its goals. Certain levels of conflict within an organization can have positive consequences for individuals as well as groups and for the organization as a whole, giving people the opportunity to identify issues and opportunities that go unnoticed. It is one of the most challenging tasks for managers to know how to manage conflicts because the outcomes should be positive for both the organization and the employees. The survey concludes that people respond differently to conflicts. Most respondents try to solve the problem by talking, they often keep their opinions to themselves in order to avoid conflict at all, rarely involve their superior in the conflict, etc., which ultimately leads to the conclusion that most people respond positively to conflicts which are resolved by conversation and compromise. This is desirable at the personal level, but can be crucial at the organizational level.

4 REFERENCES

- [1] Buble, M. (2011). *Poslovno vođenje*. Zagreb. M. E. P
- [2] Srića, V. (2004). *Biblija modernog vođe*. Zagreb. Znanje
- [3] Karlić, T. & Hadelan, L. (2011.) Stil vođenja u funkciji uspješnog poslovanja tvrtke. *Praktični menadžment*, 2(3), 67-72. Retrieved from <https://hrcak.srce.hr/76455>
- [4] Polić, I. (2019), Leadership styles determine proactivity of employees - A vessel's example. *Pomorski zbornik*, 56(2019), 39-58.
- [5] Samanta, I. & Lamprakis, A. (2018) Modern leadership types and outcomes: the case of Greek public sector. *Management*, 23(1), 173-191. <https://doi.org/10.30924/mjcmi/2018.23.1.173>
- [6] Miloloža, I. (2018) Analysis of the leadership style in relation to the characteristics of Croatian enterprises. *Interdisciplinary Description of Complex Systems*, 16(2), 249-264. <https://doi.org/10.7906/indec.16.2.5>
- [7] Garača, N. (2014) 1, 2, 3, 4... *Menadžment*. Virovitica. Visoka škola za menadžment u turizmu i informatici u Virovitici.
- [8] Sherbakov, I. D. (2014). Conflict in organizations: their causes and resolution. *Economics and management of innovation technologies*, 30(3). <http://ekonomika.snauka.ru/2014/03/4610>
- [9] Rijavec, M. & Miljković, D. (2002.) *Kako rješavati konflikte*. Zagreb. IEP
- [10] Certo, S. & Certo T. (2009). *Moderni menadžment*. 10th edition. Zagreb. MATE
- [11] Dessler, G. (2015). *Upravljanje ljudskim potencijalima*. Zagreb. MATE
- [12] Karbowski, A. & Prokop, J. (2018) R&D activities of enterprises, product market leadership, and collusion. *Zbornik radova Ekonomskog fakulteta Rijeka*, 36(2), 735-753.
- [13] Putarek, V. & Vlahović-Štetić, V. (2019) Metacognitive Feelings and Illusion of Linearity. *Psychological Topics*, 28(1), 171-192. <https://doi.org/10.31820/pt.28.1.9>
- [14] Shafique, I., Kalyar, M. N., & Ahmad, B. (2018) The nexus of ethical leadership, job performance, and turnover intention: the mediating role of job satisfaction. *Interdisciplinary Description of Complex Systems*, 16(1), 71-87. <https://doi.org/10.7906/indec.16.1.53>
- [15] Erceg, A. (2019). Information security: threat from employees, *Tehnički glasnik*, 13(2), 123-128. <https://doi.org/10.31803/tg-20180717222848>

Authors' contacts:

Jerko GLAVAŠ, PhD, Associate Professor
Faculty of Economics in Osijek,
Josip Juraj Strossmayer University in Osijek,
Trg Ljudevita Gaja 7, 31000 Osijek, Croatia
Tel: 00 385 31 22 44 00, E-mail: jerko.glavas@efos.hr

Veljko KONDIĆ, mag. ing. mech.
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
E-mail: veljko.kondic@unin.hr

Iva BALIĆ, student
Faculty of Economics in Osijek,
Josip Juraj Strossmayer University in Osijek,
Trg Ljudevita Gaja 7, Osijek, Croatia
E-mail: ivbalic@efos.hr

ARTICLE TITLE ONLY IN ENGLISH (Style: Arial Narrow, Bold, 14pt)

Ivan HORVAT, Thomas JOHNSON (Style: Arial Narrow, Bold, 11pt)

Abstract: Article abstract contains maximum of 150 words and is written in the language of the article. The abstract should reflect the content of the article as precisely as possible. TECHNICAL JOURNAL is a trade journal that publishes scientific and professional papers from the domain(s) of mechanical engineering, electrical engineering, civil engineering, multimedia, logistics, etc., and their boundary areas. This document must be used as the template for writing articles so that all the articles have the same layout. (Style: Arial Narrow, 8pt)

Keywords: keywords in alphabetical order (5-6 key words). Keywords are generally taken from the article title and/or from the abstract. (Style: Arial Narrow, 8pt)

1 ARTICLE DESIGN

(Style: Arial Narrow, Bold, 10pt)

(Tab 6 mm) The article is written in Latin script and Greek symbols can be used for labelling. The length of the article is limited to eight pages of international paper size of Letter (in accordance with the template with all the tables and figures included). When formatting the text the syllabification option is not to be used.

1.1 General Guidelines

(Style: Arial Narrow, 10pt, Bold, Align Left)

The document format is Letter with margins in accordance with the template. A two column layout is used with the column spacing of 10 mm. The running text is written in Times New Roman with single line spacing, font size 10 pt, alignment justified.

Article title must clearly reflect the issues covered by the article (it should not contain more than 15 words).

Body of the text is divided into chapters and the chapters are divided into subchapters, if needed. Chapters are numbered with Arabic numerals (followed by a period). Subchapters, as a part of a chapter, are marked with two Arabic numerals i.e. 1.1, 1.2, 1.3, etc. Subchapters can be divided into even smaller units that are marked with three Arabic numerals i.e. 1.1.1, 1.1.2, etc. Further divisions are not to be made.

Titles of chapters are written in capital letters (uppercase) and are aligned in the centre. The titles of subchapters (and smaller units) are written in small letters (lowercase) and are aligned left. If the text in the title of the subchapter is longer than one line, no hanging indents.

Typographical symbols (bullets), which are being used for marking an item in a list or for enumeration, are placed at a beginning of a line. There is a spacing of 10pt following the last item:

- Item 1
- Item 2
- Item 3

10pt

The same rule is valid when items are numbered in a list:

1. Item 1
2. Item 2
3. Item 3

10pt

1.2 Formatting of Pictures, Tables and Equations

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

Figures (drawings, diagrams, photographs) that are part of the content are embedded into the article and aligned in the centre. In order for the figure to always be in the same position in relation to the text, the following settings should be defined when importing it: text wrapping / in line with text.

Pictures must be formatted for graphic reproduction with minimal resolution of 300 dpi. Pictures downloaded from the internet in ratio 1:1 are not suitable for print reproduction because of unsatisfying quality.

10pt

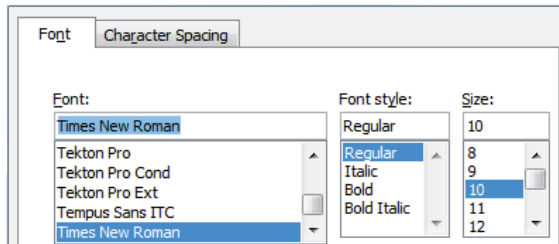


Figure 1 Text under the figure [1]

(Style: Arial Narrow, 8pt, Align Centre)

10pt

The journal is printed in black ink and the figures have to be prepared accordingly so that bright tones are printed in a satisfactory manner and are readable. Figures are to be in colour for the purpose of digital format publishing. Figures in the article are numbered with Arabic numerals (followed by a period).

Text and other data in tables are formatted - Times New Roman, 8pt, Normal, Align Center.

When describing figures and tables, physical units and their factors are written in italics with Latin or Greek letters,

while the measuring values and numbers are written upright.

10pt

Table 1 Table title aligned centre
(Style: Arial Narrow, 8pt, Align Centre)

	1	2	3	4	5	6
ABC	ab	ab	ab	ab	ab	ab
DEF	cd	cd	cd	cd	cd	cd
GHI	ef	ef	ef	ef	ef	ef

10 pt

Equations in the text are numbered with Arabic numerals inside the round brackets on the right side of the text. Inside the text they are referred to with equation number inside the round brackets i.e. “... from Eq. (5) follows ...” (Create equations with MathType Equation Editor - some examples are given below).

10pt

$$F_{\text{avg}}(t, t_0) = \frac{1}{t} \int_{t_0}^{t_0+t} F(q(\tau), p(\tau)) d\tau, \quad (1)$$

10pt

$$\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cdot \cos \frac{\alpha - \beta}{2}. \quad (2)$$

10pt

Variables that are used in equations and also in the text or tables of the article are formatted as *italics* in the same font size as the text.

10pt

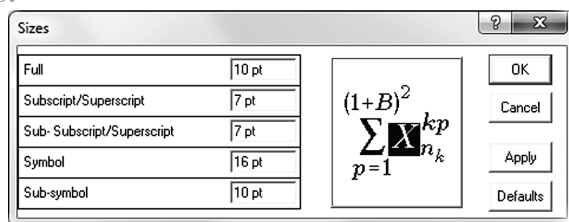


Figure 2 The texts under figures
(Style: Arial Narrow, 8pt, Align Centre)

10pt

Figures and tables that are a part of the article have to be mentioned inside the text and then connected to the content i.e. „... as shown in Fig. 1...” or „data from Tab. 1...” and similar.

10pt

2 PRELIMINARY ANNOTATION

10pt

Article that is offered for publication cannot be published beforehand, be it in the same or similar form, and it cannot be offered at the same time to a different journal. Author or authors are solely responsible for the content of the article and the authenticity of information and statements written in the article.

Articles that are accepted for publishing are classified into four categories: original scientific papers, preliminary communications, subject reviews and professional papers.

Original scientific papers are articles that according to the reviewer and the editorial board contain original theoretical or practical results of research. These articles need to be written in such a way that based on the information given, the experiment can be repeated and the

results described can be achieved together with the author's observations, theoretical statements or measurements.

Preliminary communication contains one or more pieces of new scientific information, but without details that allow recollection as in original scientific papers. Preliminary communication can give results of an experimental research, results of a shorter research or research in progress that is deemed useful for publishing.

Subject review contains a complete depiction of conditions and tendencies of a specific domain of theory, technology or application. Articles in this category have an overview character with a critical review and evaluation. Cited literature must be complete enough to allow a good insight and comprehension of the depicted domain.

Professional paper can contain a description of an original solution to a device, assembly or instrument, depiction of important practical solutions, and similar. The article need not be related to the original research, but it should contain a contribution to an application of known scientific results and their adaptation to practical needs, so it presents a contribution to spreading knowledge, etc.

Outside the mentioned categorization, the Editorial board of the journal will publish articles of interesting content in a special column. These articles provide descriptions of practical implementation and solutions from the area of production, experiences from device application, and similar.

10pt

3 WRITING AN ARTICLE

10pt

Article is written in the English language and the terminology and the measurement system should be adjusted to legal regulations, standards (ISO 80 000 series) and the SI international system of units. The article should be written in third person.

Introduction contains the depiction of the problem and an account of important results that come from the articles that are listed in the cited literature.

Main section of the article can be divided into several parts or chapters. Mathematical statements that obstruct the reading of the article should be avoided. Mathematical statements that cannot be avoided can be written as one or more addendums, when needed. It is recommended to use an example when an experiment procedure, the use of the work in a concrete situation or an algorithm of the suggested method must be illustrated. In general, an analysis should be experimentally confirmed.

Conclusion is a part of the article where the results are being given and efficiency of the procedure used is emphasized. Possible procedure and domain constraints where the obtained results can be applied should be emphasized.

10pt

4 RECAPITULATION ANNOTATION

10pt

In order for the articles to be formatted in the same manner as in this template, this document is recommended for use when writing the article. Finished articles written in MS Word for Windows and formatted according to this

template must be submitted using our The Paper Submission Tool (PST) (<https://tehnickiglasnik.unin.hr/authors.php>) or eventually sent to the Editorial board of the Technical Journal to the following e-mail address: tehnickiglasnik@unin.hr

The editorial board reserves the right to minor redaction corrections of the article within the framework of prepress procedures. Articles that in any way do not follow these authors' instructions will be returned to the author by the editorial board. Should any questions arise, the editorial board contacts only the first author and accepts only the reflections given by the first author.

10pt

5 REFERENCES (According to APA)

10pt

The literature is cited in the order it is used in the article. Individual references from the listed literature inside the text are addressed with the corresponding number inside square brackets i.e. "... in [7] is shown ...". If the literature references are web links, the hyperlink is to be removed as shown with the reference number 8. Also, the hyperlinks from the e-mail addresses of the authors are to be removed. In the literature list, each unit is marked with a number and listed according to the following examples (omit the subtitles over the references – they are here only to show possible types of references):

9pt

- [1] See <http://www.bibme.org/citation-guide/apa/>
- [2] See http://sites.umuc.edu/library/libhow/apa_examples.cfm
- [3] (Style: Times New Roman, 9pt, according to APA)
- [4] Amidzic, O., Riehle, H. J., & Elbert, T. (2006). Toward a psychophysiology of expertise: Focal magnetic gamma bursts as a signature of memory chunks and the aptitude of chess players. *Journal of Psychophysiology*, 20(4), 253-258. <https://doi.org/10.1027/0269-8803.20.4.253>
- [5] Reitzes, D. C. & Mutran, E. J. (2004). The transition to retirement: Stages and factors that influence retirement adjustment. *International Journal of Aging and Human Development*, 59(1), 63-84. Retrieved from <http://www.baywood.com/journals/PreviewJournals.asp?Id=0091-4150>
- [6] Jans, N. (1993). *The last light breaking: Life among Alaska's Inupiat Eskimos*. Anchorage, AK: Alaska Northwest Books.
- [7] Miller, J., & Smith, T. (Eds.). (1996). *Cape Cod stories: Tales from Cape Cod, Nantucket, and Martha's Vineyard*. San Francisco, CA: Chronicle Books.
- [8] Chaffe-Stengel, P. & Stengel, D. (2012). *Working with sample data: Exploration and inference*. <https://doi.org/10.4128/9781606492147>
- [9] Freitas, N. (2015, January 6). People around the world are voluntarily submitting to China's Great Firewall. Why? Retrieved from http://www.slate.com/blogs/future_tense/2015/01/06/tencent_s_wechat_worldwide_internet_users_are_voluntarily_submitting_to.html

(Style: Times New Roman, 9pt, according to APA)

10pt

10pt

Authors' contacts:

8pt

Full Name, title

Institution, company

Address

Tel./Fax, e-mail

8pt

Full Name, title

Institution, company

Address

Tel./Fax, e-mail



ichd-home.com

ICHHD 2020

14th
International
Conference on
Hydrodynamics

Hosting Institutes
CNR-INM
Roma Tre University

Rome
Aug 30 - Sept 4
2020

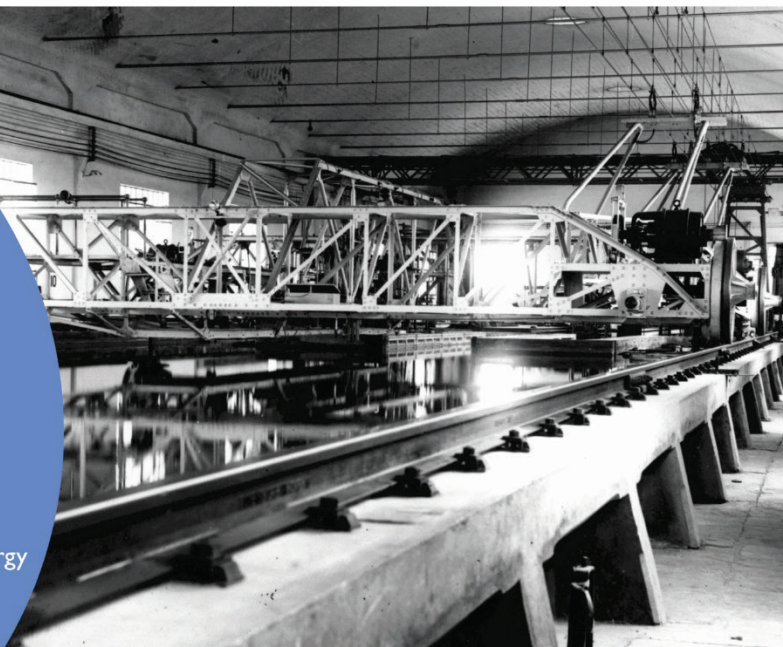
Chairmen

Prof. Massimo Gennaretti
(Roma Tre University)
Dr. Claudio Lugni
(CNR-INM)

To encourage the interdisciplinary research, each day of the Symposium, a plenary session will be dedicated to one of the following fundamental topic:

- Climate Change
- Ocean Pollution
- Marine Biotechnology
- Next Generation of Marine Renewable Energy

For each of them, a keynote lecture, followed by 2-3 extended presentations and an open discussion, will provide a multidisciplinary vision of the subject



MOTSP 2020

11th International Scientific Conference
Industry 5.0 – Return to Humanity
Management of Technology Step to Sustainable Production

June 3 - 5, 2020, Bol, island Brač, Croatia

International Conference Management of Technology – Step to Sustainable Production (MOTSP 2020), will take place from 3rd–5th June 2020 in Bol, island Brač, Croatia as a joint project organized by the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatian Association for PLM, Warsaw University of Technology–Faculty of Transport and University of North, Koprivnica. The main objective of this Conference is to gather international experts from academic entities, research laboratories and industries related to the field of Management of Technology and Sustainable Production. The Conference will also provide a platform for sharing knowledge, ideas and results between science and industry.

The management of technology, stimulation of innovation and invention and transfer of technology are important challenges of the developed countries and countries in transition.

ORGANIZATION

- Faculty of Mechanical Engineering and Naval Architecture
- University of Zagreb, Zagreb, Croatia
- Croatian Association for PLM, Zagreb, Croatia
- University of North, Koprivnica, Croatia
- IRI – Institute for Innovation and Development, University of Ljubljana, Slovenia
- Faculty for Transport, Warsaw University, Poland

MAIN TOPICS

- Management of Technology
- Production, Operation Management
- Strategic, Engineering Management
- Industrial Engineering
- Rapid Prototyping and Manufacturing
- Computer Integrated Manufacturing CAD, CAM, CAPP, CAQ
- Intelligent Manufacturing
- Product Lifecycle Management (PLM)
- Total Cost Assessment
- Ergonomics
- Sustainable Production
- Social Responsibility
- Total Cost Assessment
- Financial Sustainability
- Sustainable Transport Management
- Transport Technology and Organization



All accepted papers would be published in the Technical Journal. Paper publishing would be included by conference fee.



IMPORTANT DATES

- Deadline for abstract submission: **December 20, 2019**
Submission of paper titles and abstracts (up to 150 words) using abstract submission form at motsp.eu
- Notification of acceptance of abstracts: **January 2, 2020**
- Final submission of abstracts and full papers: **March 1, 2020**
- Notification of acceptance of full papers: **April 30, 2020**

Chairman:

Prof. Predrag Čosić, PhD, FAMENA, Department of Industrial Engineering, predrag.cosic@motsp.eu; pcosic3@gmail.com

Secretary:

Assistant Professor Tihomir Opetuk, tihomir.opetuk@motsp.eu
Assistant Professor Zdenka Keran, zdenka.keran@motsp.eu

E-mail: info@motsp.eu

Url: www.motsp.eu

INFORMATION FOR PARTICIPANTS

Conference venue:

Bluesun, hotel Elaphusa, Put Zlatnog rata 46, Island Brač, Croatia
Url: <https://www.hotelelaphusabrac.com/en/>
Tel: +385 1 3844 177; Fax: +385 1 3844 177

Participants and accompanying persons of MOTSP 2020 have a special offer from the Hotel Elaphusa 4*, Hotel Borak 3* and Hotel Bonaca 3*, Bol, island Brač.

MOTSP2020 suggest to make a reservation of accommodation ONLY through PROMO KOD on site www.blusunhotels.com.

Conference language: English

Presentation:

Presentations should take 15 minutes. Authors will have the possibility of using computers for the presentations (Power Point).

REGISTRATION

Registration should be acknowledged at the reception desk on Bluesun, hotel Elaphusa, Bol

- June 2, 2020 18:00 – 20:00
- June 3, 2020 08:00 – 09:00

Conference Fees:

Early registration fee until April 15

– EUR 460

Late registration fee

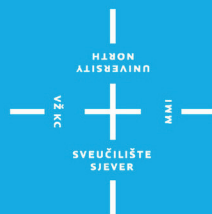
– EUR 490

Deadline for registration fee payment:

– May 15, 2020

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 13 – BROJ / NUMBER 4

PROSINAC 2019 / DECEMBER 2019 – STRANICA / PAGES 261-362



**Sveučilište
Sjever**

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TEHNICKIGLASNIK@UNIN.HR – [HTTP://TEHNICKIGLASNIK.UNIN.HR](http://tehnickiglasnik.unin.hr)